1-1-2016

Essays on Social Media, Hiring Networks and Firm Performance

Fujie Jin
University of Pennsylvania, jinfujie@wharton.upenn.edu

Essays on Social Media, Hiring Networks and Firm Performance

Abstract
This dissertation includes three essays that examine the impact of information technology on organizational performance. In the first essay, we examine the impact of network structure in the hiring of IT versus non-IT labor on firm performance. We find that hiring IT workers from a structurally-diverse network of firms has a positive effect on firm productivity, while there is no similar effect for non-IT labor in general. We attribute this to the different nature of knowledge diffused through the two types of labor: IT labor enables the transfer of new and innovative firm practices which benefits from diversity, while non-IT labor flows are associated with implementation of organizational practices, which may benefit from hiring more employees with a common knowledge base.

In the second essay, we examine the economic value of social media investments and identify the organizational complements. We argue that social media brings in large amounts of real-time data, requiring a sufficient amount of data analytical skills for organizations to effectively process the information and integrate it into organizational decision making. We find evidence that the value of social media investments is higher in firms with a larger pool of data analytic skills in the labor force. In addition, social media's positive impact on firm performance extends beyond the marketing department, and is further increased when the data analytic skills are dispersed throughout the firm.

In the third essay, we investigate whether startup firms' use of social media is associated with increased success in raising venture capital. We find that an active social media presence and strong Twitter influence increase a startup's chances of receiving more funding and from a larger pool of investors. Specifically, social media improves startup funding success through two channels: reducing the search cost for investors to discover new investment opportunities and providing an additional channel of information for investors to better evaluate startup quality.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Operations & Information Management

First Advisor
Lorin M. Hitt

Second Advisor
Lynn Wu

Keywords
Complementarity, Data Analytics, Network, Productivity, Social Media, Venture Financing

Subject Categories
Databases and Information Systems | Economics | Entrepreneurial and Small Business Operations

ESSAYS ON SOCIAL MEDIA, HIRING NETWORKS AND FIRM PERFORMANCE

Fujie Jin

A DISSERTATION

in

Operations, Information and Decisions

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

____

Lorin M. Hitt
Zhang Jindong Professor; Professor of
Operations, Information and Decisions

Co-Supervisor of Dissertation

____

Lynn Wu
Assistant Professor of Operations,
Information and Decisions

Graduate Group Chairperson

____

Eric Bradlow, K.P. Chao Professor; Professor of Marketing, Statistics, and Education

Dissertation Committee:

Lorin M. Hitt, Zhang Jindong Professor of Operations, Information and Decisions
Lynn Wu, Assistant Professor of Operations, Information and Decisions
Kartik Hosanagar, Professor of Operations, Information and Decisions

ESSAYS ON SOCIAL MEDIA, HIRING NETWORKS AND FIRM PERFORMANCE

ACKNOWLEDGMENT

First and foremost, my sincere thanks to Lorin Hitt and Lynn Wu: I am so blessed to have you as my advisors. It probably takes the volume of another dissertation to thank you properly for the guidance, teaching and support all these time.

I would like to thank Kartik Hosanagar for being a part of my committee and for his great insight, advice and suggestions. Special thanks to Xuanming Su and Noah Gans for the supportive supervision through the PhD program. Thanks to Andy Wu for being such a great co-author. I am also indebted to various faculty members, staff and PhD students at Wharton, who provided valuable input and help of various types that facilitated the completion of this dissertation.

Finally, I want to thank my parents, for their unconditional love and for always having faith in me.

ABSTRACT

ESSAYS ON SOCIAL MEDIA, HIRING NETWORKS AND FIRM PERFORMANCE

Fujie Jin

Lorin M. Hitt

Lynn Wu

This dissertation includes three essays that examine the impact of information technology on organizational performance. In the first essay, we examine the impact of network structure in the hiring of IT versus non-IT labor on firm performance. We find that hiring IT workers from a structurally-diverse network of firms has a positive effect on firm productivity, while there is no similar effect for non-IT labor in general. We attribute this to the different nature of knowledge diffused through the two types of labor: IT labor enables the transfer of new and innovative firm practices which benefits from diversity, while non-IT labor flows are associated with implementation of organizational practices, which may benefit from hiring more employees with a common knowledge base.

In the second essay, we examine the economic value of social media investments and identify the organizational complements. We argue that social media brings in large amounts of real-time data, requiring a sufficient amount of data analytical skills for organizations to effectively process the information and integrate it into organizational decision making. We find evidence that the value of social media investments is higher in firms with a larger pool of data analytic skills in the labor force. In addition, social media's positive impact on firm performance extends beyond the marketing department, and is further increased when the data analytic skills are dispersed throughout the firm.

In the third essay, we investigate whether startup firms' use of social media is associated with increased success in raising venture capital. We find that an active social media presence and strong Twitter influence increase a startup's chances of receiving more funding and from a larger pool of investors. Specifically, social media improves startup funding success through two channels: reducing the search cost for investors to discover new investment opportunities and providing an additional channel of information for investors to better evaluate startup quality.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1 : Are All Spillovers Created Equal?

A Network Perspective on IT Labor Movements

Abstract

This study examines how characteristics of inter-firm labor-flow networks affect firm productivity. Using employee job histories to trace labor movement between organizations, we construct labor-flow networks for both IT-labor and non-IT labor and analyze how a firm's network structure for the two types of labor affects its performance. We find that hiring IT workers from a structurally-diverse network of firms can substantially improve firm productivity, but the same is not true for non-IT labor where we find little benefit of network diversity. We hypothesize that these results reflect differences in the types of knowledge diffusion facilitated by different types of labor flows, with IT labor enabling the transfer of new and innovative firm practices which benefits from diversity, while non-IT labor flows are more closely associated with implementation of complementary organizational practices, which may benefit from a critical mass of workers with a common knowledge base. Together, these results demonstrate the importance of incorporating a network perspective in understanding the full impact of spillover effects from organizational hiring activities.

1.1 Introduction

Technology innovation can generate productivity spillovers as technology use and know-how propagate from earlier adopters to later ones (Bartelsman, Caballero and Lyons 1994, David 1990, Dedrick, Gurbaxani and Kraemer 2003, Tambe and Hitt 2013). Access to superior technology is a primary driver for firms' IT productivity, and spillovers created by the mobility of IT workers can be an important part of technology acquisition, especially for process knowledge that is required to exploit the benefits of IT hardware and software (Tambe and Hitt 2013). It has been observed that IT workers have high labor mobility, and IT practitioners have noted that one way these workers obtain career advancement is through acquiring skills at one employer which can be transferred to future employers (Dedrick et al. 2003, Draca, Sadun and Van Reenen 2006). As IT workers move from one company to another, they also diffuse knowledge about IT implementations and the associated business process innovations across firm boundaries. Such a source of spillovers can have significant implications for firm productivity and long-term economic growth.

There exists a substantial body of prior work tying firm- and economy-wide productivity gains to knowledge flows among firms either through the mobility of R&D workers (Almeida and Kogut 1999, Audretsch and Feldman 1996, Jaffe, Trajtenberg and Hennderson 1992), through buyer-supplier interactions (Bartelsman et al. 1994) or through collaboration networks (Powell et al. 1999, Singh et al. 2015, Singh 2005). A more recent stream of literature has begun to explore the presence of inter-firm IT spillovers between related industries (Cheng and Nault 2012, Tambe and Hitt 2013) and technology-mediated communities (Huang et al. 2013). We contribute to this stream of research by applying network analysis to better understand how the structural characteristics of firm-level labor flow networks affect firms' abilities to capture the spillovers, and how labor type can mediate the effect of the structural characteristics. In particular, we are especially interested in whether network diversity – the variety of firms from which a focal firm hires new employees – affects the spillover benefits firms receive from worker mobility. To the extent that a diverse hiring network exposes a firm to a greater variety of potential information, network diversity could increase spillover strength. However, the opposite can be true if repeated hiring from a similar knowledge base is useful in reinforcing firm-wide business practices. Indeed the relationship can be different for different types of workers depending on whether they are engaged in innovative or exploitative activities or whether the information is more easily transferrable among individuals. Our approach is therefore to examine whether the spillover benefits of IT labor flows are influenced by network diversity, and to contrast these effects with the spillover benefits of non-IT labor.

To illustrate the importance of incorporating network positions in the productivity analysis, consider the following example: Firms A and B hire the same number of new employees each year from other firms with the same average IT investment per worker. They would thus enjoy the same level of IT spillovers, according to measures used in the previous literature. However, consider the case in which Firm A hires only from a cohesive network consisting of firms that all mutually hire from each other, while Firm B hires from a variety of sources that do not hire from each other. To the extent that there exists inter-firm variation in IT practices, Firm B is more likely to get exposure to the latest technological innovations and the best implementation practices associated with the technology. Therefore, Firm B has a greater possibility of identifying new innovations through the combination of best practices in different technological areas. However, if transferring innovations from external firms requires a substantial number of similarly experienced employees, then Firm A may have an advantage. Firm A will be able to utilize the incoming knowledge more successfully, even if the total amount of new information inflow is lower. Therefore, the network positions in firms' labor-flow networks have the potential to influence the ability to obtain and utilize different types of external knowledge.

The existing social networks literature argues that network effects arise from network structures, beyond the properties of the immediate neighbors. In particular, triadic relationships ("friends of friends") are shown to play an outsized role in improving various innovation and performance outcomes in many industries and settings (Aral, Brynjolfsson and Van Alstyne 2012, Burt 1992, Cross and Cummings 2004, Oestreicher-Singer and Sundararajan 2012, Uzzi and Gillespie 2002, Wu 2013). Extending a similar logic to the hiring network of firms, we should expect firms to obtain spillovers if they hire from IT-intensive firms (the equivalent of direct effect from immediate neighbors) and also if they hire from other firms located in different parts of the inter-organizational network (the equivalent of "triadic relationships"). By using a network perspective for studying IT spillovers among firms, we bridge the social network literature with the IT productivity literature to more comprehensively understand how firms benefit from spillovers generated and transmitted through each other's IT investments and skilled labor.

Using a novel data set of individual employment histories (resumes), we are able to tie detailed employee movement data to specific firms, overcoming data limitations that constrained most previous works on labor mobility. This database contains information on several hundred thousand US-based IT workers and millions of non-IT workers. Matching the career histories of these individuals to the firm-level performance data of the employers, we tracked job movements of as much as one-sixth of the total workforce in the US over the last twenty years. From this rich data set, we are able to obtain precise measures for IT-related and non-IT related productivity spillovers, as well as measures for firms' positions in the inter-firm labor networks formed by IT and non-IT labor flows. Other information provided by the database, including educational levels, prior work experience and geographic location, can be utilized to further distinguish spillovers arising from variations in incoming labor quality.

A challenge in interpreting this type of analysis is that there may be potential endogeneity in both the factors that generally affect productivity as well as endogeneity in the quantity and pattern of labor flows across different firms and over time. To address general unobserved productivity shocks in the measurement framework, we use the Arellano-Bond estimator, which has proven effective for the study of IT and R&D productivity (Levinsohn and Petrin 2003, Olley and Pakes 1996, Tambe and Hitt 2012). Performance-related effects of network relationships are often confounded with unobserved similarity of entities that participate in the same network. To address this problem, we focus on specific network path relationships (worker flows) that may be less susceptible to this bias than more general industry-related network measures (Tambe and Hitt 2012). In addition, we use instrumental variables leveraging the fact that implementation of large scale enterprise systems provide a substantial shock to the structure of hiring relationships and the demand for different types of employees skills. This shock can then provide a source of variation for network position for both IT workers (those involved in the implementation) as well as non-IT workers who will ultimately be responsible for implementing complementary business practices and utilizing these systems. Specifically, we focus on the implementation of ERP systems because these implementation events are identifiable in the data and they are typically substantial enough to have firm-wide implications for human resource requirements and firm operations (Aral, Brynjolfsson and Wu 2012). As firms scale up their capabilities to implement ERP they will likely shift the types of skills they need, and also provide a different skill base to the neighboring firm for subsequent hires. Thus, we can use this technology-induced variation in networks to address the causality concerns from network structure.

Our findings suggest that the structural characteristics of a labor-flow network indeed yield significant impacts on firm productivity beyond the direct spillover generated by the immediate labor flow from other IT-intensive firms. In particular, a firm's local network diversity—or the brokerage position that a firm occupies in the hiring network—has an effect of similar size as the direct labor flow on productivity. After we further distinguish between labor flows for IT and non-IT workers, we find the network characteristics that generate the highest levels of productivity are different for the two types of labor. Whereas IT-labor flow networks benefit from having a high degree of network diversity, the same is not true for non-IT labor networks. Since the firm's network position can (and observably) differ for these types of labor their effects are not mutually exclusive – a firm can hire from a variety of firms for acquiring IT skills but concentrate their efforts on hiring other workers from only a small number of firms. These results suggest that both IT and non-IT workers generate considerable spillovers, but the mechanisms by which they do so are different. One possible explanation is that the identification of process innovations benefits from a high degree of network diversity because it broadens the potential amount of unique information. However, the implementation of these practices to end-users is more difficult and requires a greater number of employees with the requisite skill, perhaps due to the higher tacit knowledge requirements for these types of organizational changes.

1.2 Theory

The return on IT investments varies across firms, with some firms experiencing outsized returns for using IT, while others experience minimal and sometimes even negative returns (Aral et al. 2012, Bresnahan, Brynjolfsson and Hitt 2002, Lin, Lucas and Bailey 2011). One key differentiating factor is the co-adoption, inside a firm, of various IT-related business transformations associated with IT investments – firms that are able to adopt complementary organizational changes have been shown to receive higher returns on IT (Brynjolfsson and Hitt 2000, Nagle 2014, Saunders and Brynjolfsson 2013, Tafti, Mithas and Krishnan 2013). Because these technical know-how and specialized implementation skills are often embodied in the IT workforce, IT workers are in a unique position to facilitate the diffusion of these practices. Working closely with operations technology, IT workers are more likely to understand overall firm-wide business processes and discover new IT-related business innovations. As IT workers migrate to different firms throughout their careers, their labor movements can generate knowledge spillovers that substantially benefit the subsequent firms that employ them. Tambe and Hitt (2013) discuss a number of sources of IT-related spillovers created by employment mobility and document the economic significance of such spillovers. They show that the hiring of IT professionals from other IT-intensive firms is associated with an increase in performance on the order of 20-30% of the firm's own IT investment returns. Using a different perspective, Bapna et al. (2013) reinforce the importance of IT workers as a conduit for external knowledge– performance of IT workers increases with general training that could benefit future employer, but there are more limited gains from firm-specific training.

While direct spillovers from employee mobility contribute to firm productivity (Tambe and Hitt 2013), the deeper network structure surrounding the firm's hiring network can also play an instrumental role. The social networks literature has shown that the triadic structure surrounding each actor in the network can have a large effect on various measures of performance. At the individual level, the ability to bridge disconnected parties can help individuals achieve superior work outcomes, especially for workers in information-intensive industries (Aral et al. 2012, Burt 2001, Wu 2013). At the team level, externally diverse teams tend to have higher performance in research and development settings (Reagans and Zuckerman 2001). Inter-firm social network analysis also provides ample evidence that certain network structures, such as a high degree of network diversity in alliance and collaboration networks, are correlated with innovation and firm performance (Powell et al. 1999, Zaheer and Bell 2005). A leading explanation for the phenomenon is that a structurally-diverse network can provide an information advantage that is critical for work performance (Aral et al. 2012, Burt 1992, Burt 2001, Wu 2013). By reaching out to different pockets of the inter-organizational network, firms with high network diversity can access diverse and novel information which, in turn, can help firms gain competitive advantage and achieve superior results. They are more likely to be innovative, profitable and to survive industry turbulence (Burt 2004, Shipilov and Li 2008, Vasudeva, Zaheer and Hernandez 2013).

Inter-firm networks can thus have important implications for innovation and firm performance. A substantial literature on inter-firm networks has focused on alliance or collaboration networks, with links formed through patent filing and formal organizational agreements (Powell et al. 1999, Rosenkopf and Almeida 2003, Schilling and Phelps 2007). Similar effects have been observed for within-firm interpersonal communication (Reagans and McEvily 2003, Wu 2013). Currently, there exists little theoretical or empirical work on how networks that were formed through large-scale labor movements affect firm performance, and no work exists that focuses specifically on IT employment networks. Whereas the recent work by Tambe and Hitt (2013) and Huang et al. (2013) has documented the effect of direct spillovers from the IT workforce, we differ by examining how the network structure or the triadic connections surrounding the hiring firm affects the return on IT investment and the firm's overall performance.

Firms' hiring networks can play an important role in facilitating the transfer of technical skills, because much of the technical know-how accessible by a firm is in the heads of the employees, particularly for information-intensive industries. Often, such technical know-how is best acquired through on-the-job learning from previous employers. If a firm hires workers from other IT-intensive organizations, it can derive substantial benefits in labor productivity from the knowledge carried over by the new hires (Attewell

1992, Bresnahan et al. 2002). In addition to directly hiring from IT-intensive firms, a structurally-diverse hiring network that spans multiple groups of firms can also be critical for competitiveness. Especially with the fast pace of IT innovations, technical knowledge and implementation practices can quickly become obsolete as new technology innovations are constantly being discovered. Thus, having access to new technology innovation and associated business practices is essential to staying competitive. By hiring experts from different regions of the inter-firm labor flow network in a structurally-diverse network, firms can become early adopters of new technologies and successful users of the latest practices. These firms are more likely to discover new information helpful for improving productivity because information transmitted through a structurally-diverse labor flow network is more likely to be unique and novel. This type of information advantage is particularly helpful in the high tech sector that is marked by fast changing pace of innovation and uneven distribution of information among firms (Aral and Van Alstyne 2011).

In contrast, an alternative strategy to benefit from IT spillovers is to hire many employees from a concentrated few IT-intensive firms that also hire from each other, but information thus obtained is likely to be redundant because a structurally-cohesive community of firms tends to have similar information sets. The confines of such constrained networks render firms unlikely to discover new technical innovations or novel ways of using the existing technologies. Thus, although two firms both hire from the same number of other firms and all of which have the same IT investments, the two firms can still have substantially different access to information depending on where its hiring partners acquire employees. When a firm's hiring partners acquire employees from a much broader set of more distant firms, it can access a more diverse pool of knowledge than a firm with the same number of hiring partners that do not acquire employees broadly. Thus, without examining the triadic relationships in a firm's hiring network (e.g., firm A hires from firm B, which hires from firm C), it would be difficult to capture the full impact of the knowledge spillovers from hiring because the deeper network structure surrounding a hiring firm can substantially affect both the quantity and the quality of knowledge spillovers.

Hypothesis 1: *All else equal, firms with high network diversity in the IT-labor network are more productive than firms with high network cohesion in the IT-labor network.*

Whereas hiring IT labor from a diverse range of firms improves firm productivity, hiring non-IT workers may require a different configuration of the hiring network. To fully leverage new IT investments, firms often need to significantly change their existing business practices. Although it may be straightforward to change the technological infrastructure and dictate a set of new business practices for an organization, the effective adoption of these practices and the surrounding complementary changes may be more difficult. For instance, the adoption of innovations may require tacit knowledge of end users, knowledge which can only be built through direct experience or contact with other employees with similar experience (Mithas and Krishnan 2008). Simply hiring a few workers from various firms that use the best practices that could benefit the firm would be insufficient if the firm lacks the absorptive capacity to fully understand those practices (Cohen and Levinthal 1990). Transfer of knowledge across firm boundaries is an inherently challenging task (Singh 2005) and only through repeated exposures from various angles and seeing concrete ways of using these practices can a firm internalize the knowledge and appropriately use these practices (Centola and Macy 2007). Furthermore, because these organizational practices are often embedded in the larger organizational context, a firm is unlikely to benefit from the implementation of these practices if the firm does not understand their relationships to the existing practices and also to the larger organizational context. Although network diversity can provide unique and novel information, firms would not benefit from the information unless they knew how to effectively use it, particularly for complex process-oriented practices (Huang et al. 2013). Therefore, a company may need to hire a number of workers with experience in the relevant practices to overcome the natural resistance to changing to new organizational practices and to understanding how to effectively use the new practices. A cohesive group of hires can help firm internalize the information conferred, especially if the information is complex, tacit and requires repeated exposure and iterations. These observations collectively suggest that a firm is more likely to successfully implement various organizational practices by hiring non-IT workers from a structurally-cohesive labor network.

Hypothesis 2: *All else equal, firms with high network cohesion in the non IT-labor network are more productive than firms with high network diversity in the IT-labor network.*

1.3 Data and Setting

General Data Description

For the purpose of this study, we constructed a model of the inter-firm hiring network among US firms by extracting the labor movement information from an extensive database of individual resumes. The resume dataset was collected from a leading online job search website in 2007 and consists of more than 10 million full-text individual resumes. The resumes contain detailed career histories of each person including the employer, job title, and the duration of each period of employment in full text as well as in structured data fields. Users also provide other demographic and human capital information. The data set includes career information for both full-time and temporary employees, which is especially important for IT workers who are often contractors, and these employees may be more mobile and contribute a larger proportion of knowledge flow among firms. This dataset has previously been described in detail in (Tambe and Hitt 2013) and has been verified and compared against the Current Population Survey (CPS) data to ensure that the sample of the data is representative of the US labor market.

Information on employer name, dates of employment, position and job title are most relevant for the construction of an inter-firm hiring network. Consider a network graph, in which firms are represented by nodes, and firm-to-firm employee movements are represented by edges. We identify a directed edge between a focal firm and another firm if the former hired one or more employees from the latter in the previous five years.[1] The weight of the edge can either be the total number of employees following this path or some

[1] The length of this window should be long enough to capture representative labor flows between the companies, but no so long that the impact of corresponding labor flows is no longer relevant. We report the findings from using labor flow in the most recent five years. We also verified that intervals of four years or six years yield comparable results.

function of the number of employees and characteristics of the originating firm (e.g., IT intensity as in Tambe and Hitt 2013). We distinguish IT employees (about 15% of the sample) from other employees by job title[2], and we construct networks for both types of employees. In this paper, we focus on total labor flow (although we see similar results if we use the IT-intensity weighted flows in Tambe and Hitt 2013). We also conducted a sensitivity analysis to see if our model of the network is prone to discrete changes if we use different ranges of years to construct edges and found that our network model is stable. In total, we extracted 474,511 individual job movements between companies from the dataset, covering 10,207 unique source companies (where the workers move from) and 9,628 unique target companies (where the workers move to).[3] The dataset of network measurements as constructed above is then joined with financial performance measures of publicly traded companies in the Compustat dataset. The final dataset is a 20-year panel of data spanning 1987 to year 2007 for 6,442 unique companies.

Figure 1.1 illustrates an example of a hiring network, showing 70 companies in the healthcare industry for the year 2005. The nodes are color-coded from blue (low) to red (high) according to the number of neighbors (in-degree); each edge is color-coded by the weight assigned to it. We can observe firms' different hiring patterns from this illustration, including: 1) hiring a few employees from many companies that also hire from a diverse set of other companies (e.g., node A); 2) hiring many employees from only a couple of

[2] IT jobs are identified by titles that are clearly IT workers (software engineer, systems analyst, programmer analyst) or contain keywords that suggest the employee is an IT worker. Typical non-IT worker titles are: sales associate, administrative assistant, customer support representative.

[3] Note that we are only observing a portion of the labor-flow network because our graph is limited to firms that have complete complementary data.

companies that also hire from each other (e.g., node B). This difference in hiring strategies can be characterized by the firm's position in the hiring network: 1) structurally-diverse positions, having labor flows from non-redundant groups of companies (e.g., node A); 2) structurally-cohesive positions, having labor flows within a group of close-knit firms (e.g., node B). The following section describes the variables used to measure firms' network positions and the empirical strategies for identifying how the hiring-network structure influences firm performance.

Variables and Methodology

Following the classic methods used in information systems and productivity literature, (e.g., Brynjolfsson et al. 2003), we employ a Cobb-Douglas production function framework, relating firms' output ($Sales$), to primary firm level control variables, materials (M), capital (K), non-IT labor (L_nonIT), IT labor (L_IT) and additional measures of spillover effects and network diversity (*IT Pool*, *Network diversity*). We include year effects in all regressions to control for time-specific productivity shocks and firm fixed effects to address firm and industry heterogeneity.

$$logSales_{i,t} = \alpha + \beta_1 logM_{i,t} + \beta_2 logK_{i,t} + \beta_3 logL\_nonIT_{i,t} + \beta_4 logL\_IT_{i,t} + \beta_5 logIT\_Pool_{i,t} + \gamma * network\ diversity_{i,t} + firm\_controls + year\_controls + \varepsilon_{i,t}$$

The output and input measures are constructed using the same techniques as in earlier studies (Brynjolfsson and Hitt 2003, Hall 1990, Tambe and Hitt 2012, Tambe and Hitt 2013). The primary firm level controls are directly derived from Compustat. The direct spillover effect is measured using IT Pool, which is constructed the same way as Tambe and Hitt (2013). It is a weighted average of each firm's IT investment as detailed below. We also introduce a network diversity measure to capture the extent to which a firm hires from a network of firms that is structurally-diverse. The effect of a firm's network diversity on firm productivity is the focus of the paper. To understand the potential mechanism of how network diversity drives firm performance, we separate the IT labor flow network from the non-IT labor flow network and examine how the network diversity for each affects firm performance differently.

$$
\begin{aligned}
logSales_{i,t} = \alpha + \beta_1 logM_{i,t} + \beta_2 logK_{i,t} + \beta_3 logL\_nonIT_{i,t} + \beta_4 logL\_IT_{i,t} + \beta_5 logIT\_Pool_{i,t} \\
+ \gamma_1 * IT\_network\_diversity_{i,t} + \gamma_2 * nonIT\_network\_diversity_{i,t} \\
+ firm\_controls + year\_controls + \varepsilon_{i,t}
\end{aligned}
$$

Network Diversity

We adopt a common measure of network diversity that has been used extensively to study network positions of individuals and firm: Burt's (1992) Network Constraint. This measure has been widely used for capturing the extent to which an actor bridges disconnected groups. The specific computation is: $C_i = \sum_j \left(p_{ij} + \sum_q p_{iq} p_{qj}\right)^2, q \neq i, j$, where p_{ij} is the proportion of actor i's network time and energy invested in communicating with actor *j:* $p_{ij} = a_{ij} / \sum_{k \neq i} a_{ik}$, and a_{ij} measures the raw instances of communication between actor *i* and *j*. In our context, communication is measured as the labor flow between Firm i and Firm j. From here, network diversity is computed as: $Network\ Diversity = 1 - C_i$. This variable is larger if the focal firm is connected to groups of firms that are sparsely connected to each other. To provide the inputs to this measure, which require observing flows though the network, we consider a 5-year window of hiring activity. Sensitivity analysis suggests that this measure is largely stable with different window width, but does vary over time as firms adjust their network positions, and none of our results appear to be affected materially by the choice of the window width.

We chose Burt's measurement for network diversity because it best captures the structural network characteristics of the focal firm's connections and its exposure to diverse information carried over by the incoming labor. This measure has been widely used to demonstrate significant impact of the structural diversity of a network on performance in various settings. For example, Zaheer and Bell (2005) examine networks of firms and show having structural diverse networks is correlated with novel recombination of ideas and innovation. Reagans and Zuckerman (2001) have used the same measure to examine the performance of teams. They found that externally diverse (as measured using Burt's Constraint) but internally focused teams are the optimal configuration for improving team performance. In information system research, previous studies have also used this measurement to gauge information diversity as the result of having a structurally diverse network and have found that network diversity can positively increase the amount of unique information acquired by a node in the network (Aral and Van Alstyne 2011, Wu 2013).

Overall, the network diversity measure captures the ability of a node to link across disconnected groups. Two nodes could have very different network diversity even when they have the same in-degree or network size. When none of the network neighbors of a node are linked with each other, the node has much greater network diversity than a node with the same number neighbors but all its neighbors are connected with each other. High network diversity suggests exposure to diverse and novel information, from access to several pockets in the overall network through its network neighbors. This is particularly important for the information technology areas, where new advances and continuing improvements occur frequently. When firms can obtain new information quickly through their hiring network, they have substantial advantage in using the information strategically. By contrast, in a constrained network, the focal firm can essentially access information only from a single group of contacts, where the diversity and uniqueness of the information exchanged is likely to be lower. Therefore, a firm in a constrained network is less likely to obtain new and unique information quickly and is less able to take advantage of the information and advances in the IT field.

Network diversity also has an advantage for analysis because it is a local property of the network, relying on data from firms up to two network degrees separated from the focal firm. Other potential diversity metrics (such as betweenness centrality and closeness centrality) require knowing the structure of an entire network which makes them more vulnerable to missing links and measurement errors, and thus they are likely to plague any attempt (including ours) to measure the structure of large scale networks in realistic settings.

IT Pool

The external IT pool variable is constructed following the techniques in Tambe et al. (2013) to measure the direct spillover effect or the dyadic effect. This measure is calculated as the IT concentration of neighboring firms weighted by their share of IT labor inflow to the focal firm. We measure IT concentration as the ratio of the number of IT employees to the total number of employees in a firm.

$$IT\ pool\ = \sum_{i\in\{Hiring\ Sources\}} IT\ concentration_i * p_i$$

where $p_i = \frac{inflow_i}{\sum_{k\in\{Hiring\ Sources\}} inflow_k}$ and $IT\ concentration_i = \frac{IT\ employees}{total\ employees}$

Summary statistics for the main variables are reported in Table 1.1. The network diversity measures are appropriately standardized in the regressions (centered at zero with a standard deviation of 1). Correlations between the main variables are reported in Table 1.2. The correlation between network diversity measures from the IT and the non-IT hiring network stands at 0.48, which suggests that firms' hiring networks could differ depending on the labor types. Different industries, defined by 2-digit NAICS codes, are covered reasonably well by our sample, as the industry breakdown in Table 1.3 shows.

1.4 Results

We examine the impact of firm's network positions in the labor flow network on firm productivity. In Column 1, of Table 1.4, we show the result using a standard Cobb-Douglas production framework, and separating total labor into IT and non-IT labor. Overall, the estimates of most production factors are consistent with prior work on firm level production functions, including those that have been estimated for measuring IT productivity (Brynjolfsson, Hitt and Yang 2002, Hitt, Wu and Zhou 2002, Tambe and Hitt 2013). While both IT and non-IT contribute positively to firm output, the marginal product of IT labor is higher than the marginal product of non-IT labor as has been commonly observed (see Brynjolfsson and Hitt, 1996 or Tambe and Hitt, 2013 for a variety of possible explanations).

Column 2 includes the IT pool measure to capture the spillover generated by hiring IT workers from other companies. The IT pool measure of each firm is computed as the weighted average of the IT worker share of all neighboring firms, with the weight calculated as the percentage of employees hired from each firm (this is the approach used in (Tambe and Hitt 2013)). Consistent with previous studies, we show that hiring more employees from IT-intensive companies can significantly affect firm productivity. To this base specification we add network diversity (Column 3), which provides further information about the firms' overall network position. We find that network position in the labor flow network has a substantial relationship with productivity –a one-standard-deviation increase in network diversity, including both IT and non-IT hires, corresponds to a 0.90% increase in productivity. The effect of network diversity on firm productivity is in addition to the effect of direct IT spillovers which continues to have a positive and similar influence (see the estimates for IT Pool). This is consistent with the expectation in the social network literature that triadic structure of the network matters. Having high network diversity that endows a firm to access diverse information from a variety of sources, is conducive for innovation and performance.[4]

[4] We also repeated the estimation with Value Added as the dependent variable, or using log value of sales, minus the theoretical value of factor share for material times log value of material as dependent variable and obtained similar results. These estimates are shown in the appendix.

In Table 1.5, we further explore whether the effect of network diversity on productivity differs for firms' hiring networks of IT workers vs. a non-IT workers. If the access of novel and timely information is a key to improving productivity through hiring from a diverse network, we would expect the effect to be more prominent with IT labor movements. To explore this proposition, we constructed models of hiring networks for IT and non-IT labor separately and then calculated a firm-level network-diversity measure for each of these two networks. As the previous literature has established, IT workers typically have higher mobility than other professionals and are thus more likely to generate significant knowledge spillovers (Tambe and Hitt 2013). Although network diversity for the overall hiring network is positively correlated with firm performance, we show that the diversity in the IT-labor hiring network is primarily driving this effect. In Column 1, we see that each standard deviation increase in network diversity of the IT labor hiring network increases output by approximately 1%. This is on par with the effect of IT pool on firm performance, in which a one-standard-deviation increase in IT pool correlates with a 1.5% increase in firm performance[5]. On the other hand, the productivity effect from the network diversity of non-IT labor is more ambiguous. It is positive if introduced alone in the model (Column 2), but this may be because IT labor and non-IT labor network diversity measures are correlated (β=0.78). However, if we simultaneously consider the network diversity measures for both IT and non-IT labor (Column 3), the effect of IT labor network diversity remains positive and statistically significant, but the effect of non-IT labor network

[5] The changes in sample size from Table 1.4 is due to the fact that we are only using firms with non-missing observations of network diversity measures in Table 1.5; trying out the analysis on the original sample, filling in missing values with average numbers yield consistent results.

diversity is statistically insignificant even at the 0.1 level. This suggests that despite being correlated, the IT and non-IT labor network structures could differ in how they affect productivity. The network diversity for IT labor is associated with higher firm productivity while the network diversity for non-IT labor is not. These effects are in addition to the direct spillover effects.

Next, we explore the mechanism underlying Hypothesis 1—that the positive effects of a diverse IT labor network are due to providing access to a diverse and non-redundant set of information. We argue that the value of novel information in the incoming IT labor force is likely to be higher in information- intensive firms because these firms are more likely to rely on innovative IT applications. IT helps these firms better manage the information about business processes and provide standardized frameworks to transfer knowledge from one place to another (Mithas and Whitaker 2007). As a result, a finding that information-intensive firms benefit more than other firms from network diversity would be consistent with this argument. We identify information-intensive industries as the following: 1) Information; 2) Finance and Insurance; 3) Professional, Scientific, and Technical Services; 4) Health Care and Social Assistance. This classification is largely consistent with previous studies (Mithas and Krishnan 2008).

We compare the effect of network diversity for these two subgroups of firms in Columns 4 and 5 of Table 1.5. The results largely support our hypothesis that the effect of diversity in the hiring network is larger in information-intensive industries. The coefficients for the effect of IT labor hiring network diversity shown in these two columns are significantly different ($F(2, 17826)=5.92$, $p=0.0027$). The coefficients for the effects of non-IT hiring network diversity are also significantly different across the two subgroups ($F(2, 17826)=20.01$, $p=0.0001$). Thus, network diversity for IT hiring is most valuable in industries for which IT innovation is likely to be important, consistent with our hypothesis. However, we find conflicting evidence about whether hiring non-IT workers in a structurally-cohesive network is beneficial. The coefficient for the effects of non-IT hiring network diversity is statistically insignificant (see Column 3), but it is positive if the sample only includes information-intensive firms. One possibility is that network diversity continues to benefit information-intensive firms even for hiring non-IT workers because the fast-paced nature of their business requires the firm as a whole to constantly change and adapt. However, we later find that this effect also appears to be related to the variety of industries from which a firm hires–after controlling for the industry diversity of all hires, we find that the IT diversity measure is stable, but the non-IT diversity measure becomes small and statistically insignificant.

We also examined whether having a structurally diverse IT labor network can actually generate the information benefits as theorized. If our hypothesis is correct, firms that hire from structurally diverse networks should have an increase in the diversity of their IT skill base. To measure this effect, we calculated the cosine similarity between vectors of 35 sub-categories of IT skills between a current observation and a five-year lagged observation for the same firm. Consistent with our arguments, we find IT network diversity to be negatively correlated with cosine similarity, suggesting that a diverse hiring network leads to a greater change in the firms' IT skills base.

1.5 Robustness Checks

Our results are potentially susceptible to bias due to endogeneity of network position as well as more traditional reverse causality or omitted variables, potentially confounding the estimated relationship between investment and performance. To address these concerns, we conducted a number of alternative analyses by including additional controls, using instrumental variables for network positions, and using a GMM and Levinsohn-Petrin regression framework to control for endogeneity in expenditure on other production inputs.

Instrumental Variables and System GMM

Reverse causality between network position and performance presents one potential concern for our study. Instead of having network positions improve firm performance, higher-performing firms may simply attract better workers from a greater diversity of firms. Thus, occupying beneficial positions in the labor-flow network is simply a consequence of being in a high-performing firm. To address this problem, we need to find a source that introduces variations in a firm's labor-flow network without directly impacting firm productivity. We use the neighboring firms' implementation of an Enterprise Resource Planning (ERP) system to instrument for the network diversity of the focal firm. ERP systems are off-the-shelf software packages that offer a variety of functions including finance, human resources, supply-chain management, consumer relationship management and business intelligence (Aral et al. 2012, Hitt and Brynjolfsson 1996, Hitt et al. 2002). Implementation and usage of these systems usually involve changes in the operations of the firms and adjustments in their personnel (Brynjolfsson and Hitt 2000).

Such adjustments would involve both IT professionals that directly engage in working with the information systems as well as other types of employees involved in implementing complementary business process changes. Therefore, when a neighbor in a firm's hiring network implements ERP, it alters the type and quantity of labor flow into and out of the neighbor firm, which consequently induces a change in network position for the focal firm as well.

To illustrate this, we show in Figure 1.2 the impact on the network diversity of the focal firm (Firm A) when its neighbor (Firm B) implements an ERP system. Firm B now has a need to hire employees with ERP experience as well as experience working with the associated process change that often accompanies the ERP implementation, and starts to hire from new firms, such as Firm C. Not only is Firm B's hiring network more structurally diverse, at the same time, this network change also propagates to Firm A's network structure. As shown in Figure 1.2, before Firm B implemented ERP, Firm A hired from three firms (Firm 1, 3 and 4), and its network diversity score is 1.07. After Firm B implemented ERP, Firm A's network diversity has increased to 1.14 despite the fact that it still only hires from the same three firms. This example show that Firm B's EPR implementation induces a change in Firm A's network diversity even though Firm A has not changed its hiring practice. Thus, we can use a neighbor's ERP adoption event to instrument for the focal firms' network diversity.

Of course, typical endogenous factors such as homophilly or social influence could drive both neighbor's adoption and the firm's own hiring practices. We address this issue by explicitly looking only at the implementation of ERP systems as opposed to the decision to purchase ERP. A typical ERP implementation takes multiple years between the decision to buy and the actual use of the system, allowing us to separately estimate the use of ERP from the purchase decisions. While free cash flow or demand shocks might affect the timing of the decision to buy an ERP system, implementation of these systems is likely influenced more by internal organizational change processes and business requirements that are independent of financial concerns (Aral et al. 2012, Hitt and Brynjolfsson 1996, Hitt et al. 2002). While a change in the firms' hiring network may also influence productivity, the change in productivity occurs through direct spillovers which are already controlled for in the model by other measures. Thus, network variation induced through IT investments at neighboring firms should be a valid instrument for a firm's network position.

We use the two implementation events for ERP systems to instrument for network positions: implementation of the basic ERP system and implementation of specific modules on top of the basic ERP system. First, we look at individuals' descriptions of job responsibilities and experiences corresponding with their employment at different firms. If an individual mentions usage of ERP systems at a firm on his/her resume, e.g., "used ERP system to conduct inventory control," then we conservatively deduce that the firm had implemented an ERP system by the termination date of employment of this individual. A second source of data comes from a major ERP vendor's record of the dates when clients purchased and implemented one specific module. This is used to construct a 0/1 variable to represent the firm's ERP module implementation status, which is 0 before the implementation and 1 afterward.

Using ERP system implementation and the specific ERP module implementation data, we calculated two sets of instrumental variables: 1) the *total number* of neighbor firms in a focal firm's hiring network that have implemented an ERP system in each year, or $Exposure_j = \sum_{i:a_{ij}>0} ERP\_Implement_i$. 2) the *fraction* of neighbor firms that have implemented an ERP system, or $Fraction_j = \sum_{i:a_{ij}>0} ERP\_Implement_i / Indegree_j$. We construct the same set of instrumental variables for the implementation of a specific ERP module. Thus, in total, we have four instrumental variables. We also control for the in-degree of the focal firm, which is the total number of neighbors from whom the focal firm hires labor. This is a simplified version of the IT Pool measure which is weighted each in-degree by its IT labor share. Similar to IT Pool, in-degree can addresses a potential scale effect if larger firms have more neighbors and, therefore, a larger number of neighbors that implemented ERP. This set of instruments appears to have sufficient first stage power, passing standard weak instrument tests for predictive ability ($F(33, 20310) = 335.09$, $p< 0.0001$; $F(33, 20310) = 194.56$, $p< 0.0001$ for network diversity of IT and non-IT labor respectively).

In addition to the ERP related IVs, we also use the outflow network characteristics to instrument for the inflow network characteristics. Presumably, when employees leave the firm (the outflows), it should not directly affect firm productivity, and yet the inflow hiring network and outflow hiring networks are often correlated. Perhaps due to their similar characteristics such as being in a similar industry or having similar needs for talent, a firm often hires from the same set of firms to which they lose their employees. Yet, with the appropriate controls, the outflow networks should not directly affect firm productivity, because people who have left the firm cannot directly contribute to a firm's performance. Thus, outflow network characteristics could serve as instruments for inflow network characteristics. Our variable of interest is network diversity for IT and non-IT labor's the inflow networks, thus we use the corresponding network characteristics for the outflow networks as instrumental variables.

In Column 1 of Table 1.6, we report the 2SLS estimates, instrumenting the network diversity of IT and non-IT labor networks with the network diversity derived from outflow networks and ERP related instruments. Similar to the earlier results in Table 1.5, the network diversity for IT labor continues to be positive and statistically significant (β=0.516, p <.001) and the network diversity for non-IT labor is positive but not statistically significant. Interestingly, the direct spillover effect (IT pool) is no longer substantial, suggesting that network diversity may have a stronger (or perhaps more precisely measurable) relationship with productivity. Alternatively this could simply be a by-product of the increased coefficient on diversity that arises from IV estimation.

Although this 2SLS estimate is higher than the estimates before, it could be that the use of instrumental variables alleviated the measurement error problems, which would have created downward bias in our prior estimates. To explore if measurement error is indeed a significant issue, we incorporate a second measure of network diversity (*two-step reach)* which is the number of firms that the focal firm can reach within two network edges. To the extent that measurement error in this construct is largely independent of the measurement error of our network constraint measure, we can use two-step reach as an instrument to correct for the effects of measurement error (Tambe and Hitt 2012). In

Column 2, we include two-step reach as an additional instrumental variable in the GMM framework, and the results are largely similar to Column 1, indicating that measurement error could be a concern and that instrumental variables can correct the bias. We also used two-step reach as the only instrument and the estimate is still higher than the fixed effects estimates, but lower than the 2SLS model that contains all the instruments. Thus, we conclude that the apparent rise in the effect of diversity is consistent with measurement error issues that would lead our earlier estimates to be conservative, rather than some other econometric problem which could result in an indeterminate direction of bias.

Next, we use dynamic GMM to address other forms of reverse causality between production inputs and outputs that have been argued to affect production functions (e.g., sales shocks leading to greater investment), which could create biases in all coefficients. Specifically, we use the Arellano-Bond/Blundell-Bover two-step robust system GMM estimation. This procedure utilizes appropriate internal panel instruments (lagged levels and differences) to estimate differences and levels regressions and then optimally weights them, yielding an increase in efficiency relative to methods using either levels or differences alone (Arellano and Bover 1995, Blundell and Bond 1998). In Column 3 of Table 1.6, we show the estimates using the GMM framework with external instrumental variables—neighbors' ERP implementation status, two-step reach and outflow network diversity measures. We use 3-period lags[6] for the endogenous variables and the external instrumental variables constructed using ERP, outflow networks and two-step reach. The Arellano-Bond test for AR(2) in first differences cannot reject the null that there is no

[6] We tested other lags as well and found that they did not qualitatively change our results.

second-order serial correlation in the residuals of the first-differencing equation. Thus, serial correlation is not an issue in the GMM estimation. Neither the Hansen *J* statistic (over-identification test) nor the difference-in-difference Hansen test (p=0.19) rejects the null that the instruments are uncorrelated with the disturbance terms, ensuring the validity of the instruments used in the GMM estimation (Roodman 2009). The estimates in Column 3 of Table 1.6 are qualitatively similar to earlier IV estimates. Specifically, a one-standard-deviation increase in network diversity for IT labor is corresponding to a 6.86% increase in firm performance. The estimate for network diversity for non-IT labor is still statistically insignificant. We continue to find that IT pool (dyadic connections) does not affect firm performance; the estimate is negative but not statistically significant. Overall all these results underscore the importance of examining triadic network structure of a firm's hiring network. Not only does the network diversity matter for firm performance, it is even more important than the direct spillovers or the dyadic connections in affecting firm performance. Furthermore, not all network diversity matters the same way; most of the benefits from network diversity come from the mobility of IT labor as opposed to non-IT labor. A diverse hiring network for IT labor could bring new and novel information to the firm, which is essential for staying competitive in the fast-changing landscape of IT innovations. However, a structural diverse hiring network for non-IT labor would not be as beneficial as that for IT labor because understanding and implementing complex organizational practices requires repeated exposure and a sufficient number of employees with similar knowledge sets.

Finally, we use Levinsohn-Petrin estimators (Levinsohn and Petrin 2003) in Column 4 to address the potential problem arising from reverse causality between output and input demand which could lead to biased production function coefficients which, in turn, generate biases in the coefficients of interest. This technique utilizes variation in materials expenses, which are likely to adapt quickly to output shocks, to estimate the effect of reverse causality and then use those estimates to obtain consistent estimates of the production function parameters.[7] The results of these estimates are similar to the other IV estimates in magnitude and directionally consistent with the prior fixed effects results. Thus, reverse causality issues with production function estimation do not appear to be leading to an overstatement of the effects of network position on performance.

Controlling for other Network Characteristics

It is possible that hiring-network diversity should also include a time element. Perhaps in addition to hiring from diverse sources of firms at any point in time, it is also important for firms to diversify hiring sources across multiple time periods. To examine this possibility, we constructed a measure of network stability to represent whether firms continue to hire from the same sources over time or whether they switch to different sources instead. Specifically, this measure is the percentage of neighbor firms from which the focal firm continues to hire workers in the current period, amongst all neighbor firms it hired from the previous period. A high network stability measure suggests tendency to hire

---

[7] For a more detailed discussion of this approach and the underlying assumptions, please refer to Olley and Pakes (1996) and Levinsohn and Petrin (2003). The estimates were performed by the LEVPET package in STATA.

continuously from the same sources over time. We calculate the network stability measures for the IT and non-IT networks respectively.

$$Network\ Stability_{i,t} = \frac{count(same\ hiring\ sources\ in\ period\ (t-1)\ and\ t)}{count(hiring\ sources\ in\ period\ (t-1)\ )}$$

Results in Table 1.7 indicate that if we introduce a network stability variable in the model, the main results are consistent with that in Table 1.5. Specifically, when hiring IT labor, network diversity can benefit firm productivity, but the stability of the hiring network does not. Adding network stability for non-IT hires in Column 2, we find that neither network diversity nor the network stability for the non-IT labor network affects productivity significantly, and only the effect of IT-labor network diversity remains positive and statistically significant. Overall, these results suggest that the network diversity for IT labor hires is important for productivity, supporting Hypothesis 1. While we find evidence that the network diversity for non-IT labor does not have an effect on firm productivity, the evidence is not sufficient to support Hypothesis 2 that a structurally-cohesive, rather than structurally-diverse, non-IT network is more beneficial for firm productivity.

We also used an alternative measure of network position (PageRank) to examine whether our results are driven by the fact that some firms may occupy a preferential (central) position in overall the labor flow network and therefore attract different kinds of labor (e.g. higher quality), which could potentially bias our estimation. Including the PageRank measure along with our primary measures of network diversity, either with IT labor alone (Column 3) or with both IT and non-IT labor (Column 4) yield similar results to our prior analyses. IT labor network diversity is associated with higher performance while non-IT labor network diversity apparently does not. Although PageRank is positive for IT labor and negative for non-IT labor, we are reluctant to interpret these measures as PageRank is a global measure of network position and its stability is vulnerable by missing links in the overall network. Nonetheless, it does not appear that adding additional controls for network structure such as PageRank has a substantial influence on our main results.[8]

There are also a number of additional sources of potential bias that typically arise in the analyses of network data (Chandrasekhar and Lewis 2011). One well-known concern is bias due to incomplete sampling when constructing a model of the network. Given that we are generating our firm network by sampling individuals who provide information about edge weight, rather than the presence of nodes, we believe we are less susceptible to this concern. That is, a node is only omitted if it has no reported employees in our sample. We are, however, concerned that our results are affected by industry diversity or differences in labor quality, which could yield diversity effects that do not arise by the process we hypothesize.

Industry Diversity and Firm Quality

Our primary argument is that hiring-network diversity brings in novel information that is unique either within or between industries through triadic information flows. Thus, we would expect that some but not all of the network-diversity effect is due to industry diversity that a firm hires from. In order to control for such differences, we include the

[8] Due to multi-collinearity issues, we could not put network diversity, network stability and PageRank in a single model

number of industries found in the incoming labor pool. To distinguish the differences between IT labor and non-IT labor hires, we calculated separate industry counts for each. The results are reported in Table 1.8.

In Column 1 of Table 1.8 we find that industry diversity does indeed increase productivity, but our IT diversity coefficient also remains positive and significant. Interestingly, in Column 2 we find that any diversity effect we found in our original fixed effects results for non-IT labor appears entirely attributable to industry variation. To the extent that our instrumental variables analyses also filter out industry variation in the network, this may explain why the fixed effects results in Table 1.5 suggest that diversity in non-IT labor increases performance while our instrumental variables analyses detect no such effect. Thus, the analysis of industry variation appears to strengthen our hypothesized connection between network diversity and performance that depends on the category of labor.

An alternative concern is that our results are simply reflecting differences in labor quality. Because better-performing firms are more attractive to higher-skilled labor regardless of prior employment, it is possible to simultaneously yield an increase in diversity and performance. To examine this possibility, we include additional controls for the quality of the labor force at each firm, including the average experience (*Avg experience*), age (*Avg age*) and education (*University degree*). The age of the employees is deduced from the reported dates of obtaining different degrees; experience is calculated as the number of years since the individual started his/her first employment; education is calculated by first tracking the highest degree obtained by each individual, then calculating the percentage of employees with a college degree or higher for each firm. These additional controls are added to the analysis and results are reported in Table 1.9.

Results in Table 1.9 indicate that including additional controls of worker quality does not alter our main results. The diversity measures retain their signs and significance, although lower in magnitude. However, we are reluctant to offer further interpretation of these results because none of the labor quality measures appear to have a substantial direct effect (except perhaps education). Regardless, these results suggest that labor quality effects are not the primary driver of our results. We also tried using other controls of labor quality, such as the gender ratio of employees, percentage of employees that graduated from top 150 universities according to US news ranking, percentage of employees with IT major degrees, and found these controls do not significantly influence our main results. In addition to quality controls for current employees at each firm, we also tried using similar controls for the quality of firms' new hires in each period. Consistently, we found that the main results are not impacted by the inclusion of these quality measures of labor inflows. Overall, these results suggest that employee qualities are not the main drives for the relationship we find between network diversity of hiring networks and firm productivity.

1.6 Discussions and Conclusion

Our objective in this study is to advance the understanding of the productivity effects of IT spillovers by examining how the structure of a firm's IT labor-flow network affects performance. Using insights from the social networks literature, we extend prior IT labor spillover research to include the effect of firm's hiring-network diversity, which we have shown to increase the spillover effect from employees, particularly IT employees. Consistent with the prior literature, we find that the movement of IT workers among IT-intensive firms brings spillover effects and improves firm performance. In addition, we find a higher performance gain if these IT workers are hired from a structurally-diverse network. Furthermore, the IV/GMM models demonstrate that network diversity may provide a stronger signal of the strength for measuring infer-firm spillovers than the direct spillover effects. Our preferred explanation is that network diversity in hiring IT workers provides organizations with access to larger amounts of novel information and exposure to new technology and business practices. This brings productivity gains to firms, for example, through facilitating firms' strategic decisions and IT innovation. Interestingly, the advantage of hiring from a structurally diverse network does not extend to other types of workers and the effects are lower in non-information-intensive industries. This is likely due to the different nature of diffusion for various types of knowledge. Compared with *technical knowledge* that is likely to be standardized across a range of firms and therefore easier to acquire, *tacit knowledge* such as implementing organizational changes is harder to transfer without frequent interactions, repeated exposures, and the understanding of the larger organizational context. Thus, a structurally cohesive network may not help a firm achieve complex organizational changes that are necessary to realize the full benefits of IT investments. While we found evidence that network diversity in non-IT labor hiring contributes less to productivity than that in IT labor hiring, we did not have sufficient evidence to conclude whether a structurally-cohesive network is more beneficial for hiring non-IT labor. Together, these results underscore the importance of understanding the network context surrounding labor mobility. Whereas a structurally diverse network is well suited for hiring IT-labor, it may not be for hiring non-IT labor. Thus, not only should the firm take into the hiring network into account, it should also consider its relation to the type of labor they are acquiring as well as the transferability of the knowledge associated with that type of worker. For hiring IT workers, the network diversity of IT labor network can substantially affect firm productivity. As the result, firms should also consider incorporating network diversity as a part of the overall strategy for hiring IT labor. By incorporating network structure into the hiring decisions, firms stand to see further benefits in productivity from the spillovers by the new employees.

Future work could potentially explore the heterogeneity in the effect of network diversity across industries, geographic regions and between startups firms and established companies. While we show that the average effect of network diversity for IT labor networks is positive, the effect could also be heterogeneous, with firms in information-intensive industries benefit more from having high network diversity than firm in non-information intensive industries. Other industry characteristics, such as different pace of innovation or industry turbulence (Brynjolfsson et al. 2008), could also affect the return to having high network diversity and could merit further studies. In addition, geographical variation could also moderate the effect of network diversity and future work could examine to extend of which firms can overcome the geographical constraint when they deploy strategy to increase their network diversity. Lastly, while our data consists of public-traded firms in the US, it is possible that smaller organization may benefit from different hiring strategies. Future work should examine the strategies of small and medium sized firms in acquiring talent and how they differ from that of more established firms.

Our findings are also relevant, as organizations face new needs of talent acquisition brought about by new technology advancements, such as the current trend of big data. The data sample we used in this paper covers a 20-year period during which several major advances in technological innovations has occurred including the commercialization of the Internet, the dot.com bubble as well as the post bubble periods. Throughout this long time frame, we consistently observe that network diversity for IT labor hiring can positively affect firm productivity. It is likely that these findings could extend to the hiring of other workers in fields that are also characterized by fast paced changes. Currently, we are experiencing another wave of technological advances in the area of big data. Skills needed for taking advantage of the data are also changing rapidly and the demand for hiring employees with the necessary data skills is also growing. Thus, lessons learned from the network structures in IT hiring networks can also shed light into how firms can strategically acquire necessary talent in order to benefit from big data and future related technological advances.

1.7 References

Almeida P, Kogut B (1999). Localization of knowledge and the mobility of engineers in regional networks. Management Science. 45(7) 905-917.

Aral S, Brynjolfsson E, Van Alstyne M (2012). Information, technology, and information worker productivity. Information Systems Research. 23(3-part-2) 849-867.

Aral S, Brynjolfsson E, Wu L (2012). Three-way complementarities: Performance pay, human resource analytics, and information technology. Management Science. 58(5) 913-931.

Aral S, Van Alstyne M (2011). The diversity-bandwidth trade-off1. American Journal of Sociology. 117(1) 90-171.

Arellano M, Bover O (1995). Another look at the instrumental variable estimation of error-components models. Journal of Econometrics. 68(1) 29-51.

Attewell P (1992). Technology diffusion and organizational learning: The case of business computing. Organization Science. 3(1) 1-19.

Audretsch DB, Feldman MP (1996). R&D spillovers and the geography of innovation and production. The American economic review 630-640.

Bapna R, Langer N, Mehra A, Gopal R, Gupta A (2013). Human Capital Investments and Employee Performance: An Analysis of IT Services Industry. Management Science. 59(3) 641-658.

Bartelsman EJ, Caballero RJ, Lyons RK (1994). Customer-and supplier-driven externalities. The American Economic Review 1075-1084.

Blundell R, Bond S (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics. 87(1) 115-143.

Bresnahan TF, Brynjolfsson E, Hitt LM (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. The Quarterly Journal of Economics. 117(1) 339-376.

Brynjolfsson E, Hitt LM (2000). Beyond computation: Information technology, organizational transformation and business performance. The Journal of Economic Perspectives 23-48.

Brynjolfsson E, Hitt LM (2003). Computing productivity: Firm-level evidence. Review of Economics and Statistics. 85(4) 793-808.

Brynjolfsson E, Hitt LM, Yang S (2002). Intangible Assets: Computers and Organizational Capital. Brookings Papers on Economic Activity. 2002(1) 137-198.

Brynjolfsson E, McAfee A, Sorell M, Zhu F (2008). Scale without Mass: Business Process Replication and Industry Dynamics. Harvard Business School Technology & Operations Mgt. Unit Research Paper (07-016).

Burt RS (1992). Structural Holes: the Social Structure of Competition. Cambridge, MA: Harvard University Press.

Burt RS (2001). Structural Holes versus Network Closure as Social Capital. Social Capital: Theory and research 31-56.

Burt RS (2004). Structural Holes and Good Ideas. American Journal of Sociology. 110(2) 349-399.

Centola D, Macy M (2007). Complex Contagions and the Weakness of Long Ties. American Journal of Sociology. 113(3) 702-734.

Chandrasekhar A, Lewis R (2011). Econometrics of Sampled Networks. Unpublished manuscript, MIT.[422].

Cheng Z, Nault BR (2012). Relative industry concentration and customer-driven IT spillovers. Information Systems Research. 23(2) 340-355.

Cohen WM, Levinthal DA (1990). Absorptive capacity: a new perspective on learning and innovation. Administrative science quarterly 128-152.

Cross R, Cummings JN (2004). Tie and network correlates of individual performance in knowledge-intensive work. Academy of Management Journal. 47(6) 928-937.

David PA (1990). The dynamo and the computer: an historical perspective on the modern productivity paradox. The American Economic Review 355-361.

Dedrick J, Gurbaxani V, Kraemer KL (2003). Information technology and economic performance: A critical review of the empirical evidence. ACM Computing surveys. 35(1) 1-28.

Draca M, Sadun R, Van Reenen J (2006). Productivity and ICT: A Review of the Evidence.

Hall BH (1990). The manufacturing sector master file: 1959-1987. National Bureau of Economic Research.

Hitt LM, Brynjolfsson E (1996). Productivity, business profitability, and consumer surplus: three different measures of information technology value. MIS Quarterly 121-142.

Hitt LM, Wu D, Zhou X (2002). Investment in enterprise resource planning: Business impact and productivity measures. J. of Management Information Systems. 19(1) 71-98.

Huang P, Ceccagnoli M, Forman C, Wu DJ (2013). IT knowledge spillovers and productivity: Evidence from enterprise software. Available at SSRN 2243886.

Jaffe AB, Trajtenberg M, Henderson R (1992). Geographic localization of knowledge spillovers as evidenced by patent citations. National Bureau of Economic Research.

Levinsohn J, Petrin A (2003). Estimating production functions using inputs to control for unobservables. The Review of Economic Studies. 70(2) 317-341.

Lin M, Lucas HC, Bailey J (2011). 'Banking'on the Internet: Does Internet Banking Really Improve Bank Performance? Available at SSRN 1989838.

Mithas S, Krishnan MS (2008). Human capital and institutional effects in the compensation of information technology professionals in the United States. Management Science. 54(3) 415-428.

Mithas S, Whitaker J (2007). Is the world flat or spiky? Information intensity, skills, and global service disaggregation. Information Systems Research. 18(3) 237-259.

Nagle F (2014). Crowdsourced Digital Goods and Firm Productivity.

Oestreicher-Singer G, Sundararajan A (2012). The visible hand? Demand effects of recommendation networks in electronic markets. Management Science. 58(11) 1963-1981.

Olley GS, Pakes A (1996). The Dynamics of Productivity in the Telecommunications Equiptment Industry. Econometrica. 64(6) 1263-1297.

Powell WW, Koput KW, Smith-Doerr L, Owen-Smith J (1999). Network position and firm performance: Organizational returns to collaboration in the biotechnology industry. Research in the Sociology of Organizations. 16(1) 129-159.

Reagans R, McEvily B (2003). Network structure and knowledge transfer: The effects of cohesion and range. Administrative Science Quarterly. 48(2) 240-267.

Reagans R, Zuckerman EW (2001). Networks, diversity, and productivity: The social capital of corporate R&D teams. Organization Science. 12(4) 502-517.

Roodman DM (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. Stata Journal. 9(1) 86-136.

Rosenkopf L, Almeida P (2003). Overcoming local search through alliances and mobility. Management Science. 49(6) 751-766.

Saunders A, Brynjolfsson E (2013). Valuing IT-Related Intangible Assets. Available at SSRN 2344949.

Schilling MA, Phelps CC (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. Management Science. 53(7) 1113-1126.

Shipilov AV, Li SX (2008). Can you have your cake and eat it too? Structural holes' influence on status accumulation and market performance in collaborative networks. Administrative Science Quarterly. 53(1) 73-108.

Singh H, Kryscynski D, Li X, Gopal R (2015). Pipes, pools, and filters: How collaboration networks affect innovative performance. Strategic Management Journal.

Singh J (2005). Collaborative networks as determinants of knowledge diffusion patterns. Management Science. 51(5) 756-770.

Tafti A, Mithas S, Krishnan MS (2013). The effect of information technology-enabled flexibility on formation and market value of alliances. Management Science. 59(1) 207-225.

Tambe P, Hitt LM (2012). The productivity of information technology investments: New evidence from IT labor data. Information Systems Research. 23(3-part-1) 599-617.

Tambe P, Hitt LM (2013). Job hopping, information technology spillovers, and productivity growth. Management Science. 60(2) 338-355.

Tambe P, Hitt LM (2013). Job hopping, information technology spillovers, and productivity growth. Management Science.

Tambe P, Hitt LM (2013). Measuring information technology spillovers. Information Systems Research. 25(1) 53-71.

Uzzi B, Gillespie JJ (2002). Knowledge spillover in corporate financing networks: Embeddedness and the firm's debt performance. Strategic Management Journal. 23(7) 595-618

Vasudeva G, Zaheer A, Hernandez E (2013). The embeddedness of networks: Institutions, structural holes, and innovativeness in the fuel cell industry. Organization Science. 24(3) 645-663.

Wu L (2013). Social network effects on productivity and job security: Evidence from the adoption of a social networking tool. Information systems research. 24(1) 30-51.

Zaheer A, Bell GG (2005). Benefiting from network position: firm capabilities, structural holes, and performance. Strategic management journal. 26(9) 809-825.

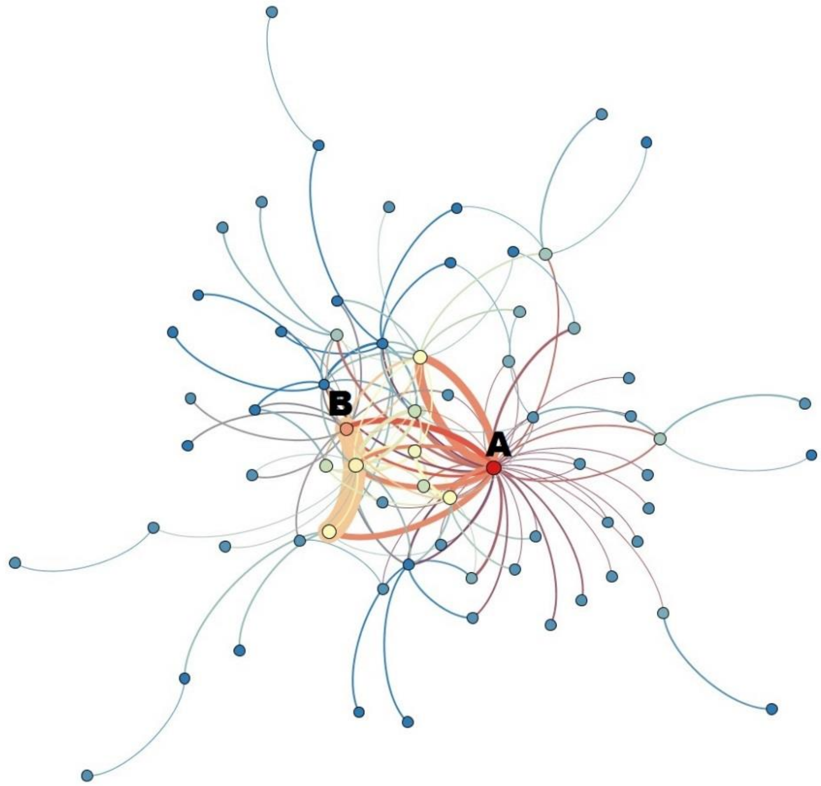


Figure 1.1. Illustration of an Inter-Firm Hiring Network

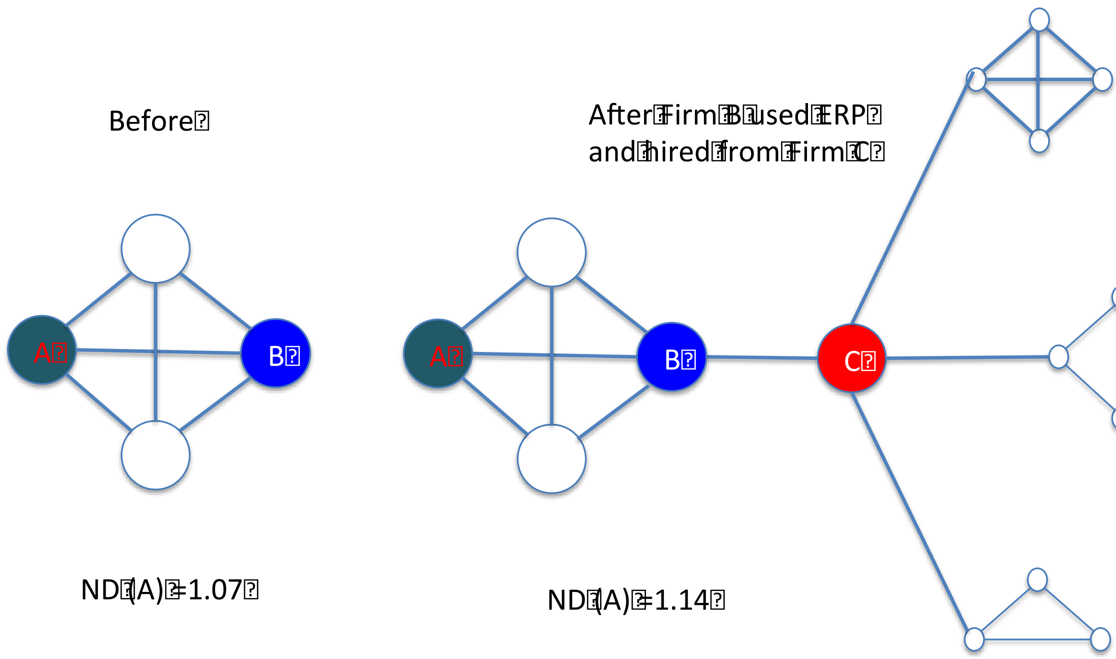


Figure 1.2. ERP Instrument Network Diversity with Neighbor's ERP Status

Notes. Graph on the left shows the hiring network of Firm A, which has a network diversity measure of 1.07. Graph on the right shows the scenario of what happened to Firm A's network diversity when Firm B, its network neighbor adopted an ERP system. Because Firm B needs to acquire workforce that can work with the ERP system, it started to hire from Firm C. As the result, Firm B's network structure has changed, but this change also affected Firm A's network. Now Firm A's network diversity has increased to 1.14 even though Firm A has not changed any of its own hiring practices. Thus, a neighbor's ERP implementation can serve as an instrument variable for focal firms' network measures.

Table 1.1. Summary Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Sales (mm)	37,713	3,790	13,507	0	337,032
Material (mm)	37,713	2,453	9,949	0.025	268,882
Capital (mm)	37,713	3,261	14,005	0.005	414,073
Non-IT Labor (m)	37,713	12.41	42.83	0.001	1,880
IT Labor	37,713	1,480	6,228	1	174,099
IT Pool	37,713	0.0882	0.112	0	1
Network Diversity	37,713	0.588	0.371	0	0.991
Network Diversity of IT	20,344	0.543	0.368	0	0.988
Network Diversity of non-IT	35,933	0.619	0.360	0	0.990

Table 1.2. Correlations of Main Variables

	1	2	3	4	5	6	7	8	9
1. Log(Sales)	1.00								
2. Log(Material)	0.96	1.00							
3. Log(Capital)	0.88	0.86	1.00						
4. Log(Non-IT Labor)	0.89	0.85	0.84	1.00					
5. Log(IT Labor)	0.67	0.63	0.63	0.64	1.00				
6. Log(IT Pool)	0.06	0.05	0.03	0.04	0.19	1.00			
7. Network Diversity	0.36	0.34	0.31	0.34	0.38	0.15	1.00		
8. Network Diversity of IT	0.39	0.37	0.33	0.36	0.49	0.28	0.53	1.00	
9. Network Diversity of non-IT	0.38	0.37	0.34	0.36	0.35	0.11	0.78	0.48	1.00

Table 1.3. Industry Composition

Industry	Freq.	Percent
Accommodation and Food Services	119	1.79
Administrative and Support and Waste Management	178	2.67
Agriculture, Forestry, Fishing and Hunt	21	0.32
Arts, Entertainment, and Recreation	33	0.50
Construction	76	1.14
Educational Services	22	0.33
Finance and Insurance	790	11.86
Health Care and Social Assistance	164	2.46
Information	949	14.24
Management of Companies and Enterprises	2	0.03
Manufacturing	2,754	41.34
Mining, Quarrying, and Oil and Gas Extraction	187	2.81
Professional, Scientific, and Technical	388	5.82
Real Estate and Rental and Leasing	98	1.47
Retail Trade	316	4.74
Transportation and Warehousing	141	2.12
Utilities	139	2.09
Wholesale Trade	228	3.42
Other Services	30	0.45
Non-classifiable	27	0.41
Total	6,662	100.00

Table 1.4. Network Positions and Productivity

DV: Log(Sales) Model	(1) FE	(2) FE	(3) FE
Log(Material)	0.618***	0.617***	0.617***
	(0.0113)	(0.0112)	(0.0112)
Log(Capital)	0.167***	0.167***	0.167***
	(0.00970)	(0.00967)	(0.00965)
Log(Non-IT Labor)	0.133***	0.133***	0.133***
	(0.00651)	(0.00651)	(0.00651)
Log(IT Labor)	0.0230***	0.0221***	0.0219***
	(0.00114)	(0.00113)	(0.00113)
Std(IT pool)		0.0181***	0.0176***
		(0.00304)	(0.00305)
Std(IT Net. Div.)			0.00904**
			(0.00418)
Constant	0.511***	0.523***	0.536***
	(0.0409)	(0.0408)	(0.0411)
Controls	Year dummies, Advertising Expense, R&D Expense dummies for missing observations		
Observations	37,713	37,713	37,713
R-squared	0.9821	0.9821	0.9823

Notes. i.Dependent variable in all regressions is log value of Sales. Regression also controls for year effect, advertising expenditure, R&D expense and dummy variables indicating where missing observations are filled in with industry average numbers; log(IT pool) is standardized for better interpretability of the coefficients;
ii. Robust clustered standard errors shown in parentheses
iii. *** p<0.01, ** p<0.05, * p<0.1

Table 1.5. Network Positions in IT and non-IT Hiring Network and Productivity

DV:Log(Sales)	(1)	(2)	(3)	(4)	(5)
Sample:	All Sample	All Sample	All Sample	Info. Intensive	Non-Info Intensive
Model	FE	FE	FE	FE	FE
Log(Material)	0.584***	0.584***	0.584***	0.452***	0.688***
	(0.0146)	(0.0145)	(0.0145)	(0.0222)	(0.0181)
Log(Capital)	0.141***	0.141***	0.140***	0.159***	0.103***
	(0.0133)	(0.0133)	(0.0133)	(0.0223)	(0.0159)
Log(Non-IT Labor)	0.210***	0.210***	0.210***	0.263***	0.171***
	(0.0150)	(0.0150)	(0.0150)	(0.0244)	(0.0183)
Log(IT Labor)	0.0236***	0.0238***	0.0236***	0.0471***	0.0159***
	(0.00200)	(0.00201)	(0.00200)	(0.00491)	(0.00202)
Std(Log(IT pool))	0.0146***	0.0157***	0.0147***	0.0226***	0.00619*
	(0.00379)	(0.00381)	(0.00380)	(0.00658)	(0.00325)
Std(IT Net. Div.)	0.0104**		0.00940**	0.0218**	0.00563
	(0.00443)		(0.00440)	(0.0103)	(0.00412)
Std(Non-IT Net. Div.)		0.00849*	0.00756	0.0264***	-0.00303
		(0.00497)	(0.00497)	(0.00948)	(0.00498)
Constant	0.328***	0.333***	0.346***	0.400***	0.283***
	(0.0664)	(0.0668)	(0.0668)	(0.108)	(0.0820)
Controls	Industry Dummies, Year Dummies, Advertising Expense, R&D Expense dummies for missing observations			Industry Dummies, Year Dummies, Advertising Expense, R&D Expense dummies for missing observations	
Observations	20,344	20,344	20,344	6,366	13,978
R-squared	0.9857	0.9857	0.9857	0.9827	0.9891

Notes. i. Dependent variable in all regressions is log value of Sales. Regression also controls for year effect, advertising expenditure, R&D expense and dummy variables indicating where missing observations are filled in with industry average numbers

ii. In (4) we use the subsample of information-intensive industries, including: Information, Finance and Insurance, Professional, Scientific and Technical Services, Health Care and Social Assistance; in (5), we use the subsample of non-information-intensive industries, which is all the remaining industries

iii. Robust clustered standard errors shown in parentheses

iv. *** p<0.01, ** p<0.05, * p<0.1

Table 1.6. GMM and Instrument Variable Estimations

DV: Log(Sales)	(1)	(2)	(3)	(4)
Model	2SLS	2SLS	GMM	LP
Log(Material)	0.631***	0.631***	0.610***	0.419***
	(0.0165)	(0.0165)	(0.0798)	(0.0983)
Log(Capital)	0.0930***	0.0931***	0.0222	0.186**
	(0.0171)	(0.0171)	(0.0624)	(0.0782)
Log(Non-IT Labor)	0.243***	0.243***	0.319***	0.214***
	(0.0269)	(0.0269)	(0.0662)	(0.0184)
Log(IT Labor)	0.0356***	0.0352***	0.0325*	0.0388***
	(0.00458)	(0.00459)	(0.0177)	(0.00307)
Log(IT pool)	0.00160	0.00119	-0.0125	0.00265
	(0.00756)	(0.00756)	(0.0762)	(0.00679)
Std(IT Net. Div.)	0.0282*	0.0304**	0.0589*	0.0358***
	(0.0150)	(0.0150)	(0.0319)	(0.00837)
Std(Non-IT Net. Div.)	0.0180	0.0201	0.0184	0.0111
	(0.0200)	(0.0202)	(0.0238)	(0.00894)
Constant	-0.287	-0.278	-1.061	
	(0.188)	(0.190)	(0.881)	
Instruments	Net. Div. outflow IT Net. Div. outflow nonIT Exposure_ERP, Exposure_HCM, Fraction_ERP, Fraction_HCM,	Net. Div. outflow IT Net. Div. outflow nonIT Exposure_ERP, Exposure_HCM, Fraction_ERP, Fraction_HCM, **Two-Step Reach**	Net. Div. outflow IT Net. Div. outflow nonIT Exposure_ERP, Exposure_HCM, Fraction_ERP, Fraction_HCM, **Two-Step Reach, 3-period lags of transformed data**	Log(Material)
Controls	Industry Dummies, Year Dummies, Advertising Expense, R&D Expense, Indegree, Dummies for missing observations			
Observations	19,126	19,126	19,126	20,344
Number of firms	2,371	2,371	2,371	2,483
R-squared	0.9462	0.9462		

Notes. i. Dependent variable in all regressions is log value of Sales. Regression also controls for year effect, advertising expenditure, R&D expense and dummy variables indicating where missing observations are filled in with industry average numbers

ii. In (1) we use 2SLS estimation, using the 4 external IVs constructed from the ERP implementation status of neighboring firms (i.e., the total number of/ the portion of neighbor firms that implemented ERP/HCM packages), and the IVs from network characteristics of the outflow network; In (2) we add two-step reach as an additional external IV in 2SLS estimation. In (3) we use Arellano-Bond System GMM estimation, specifying 3 period lags for the transformed data and 2 period lags for the differences for the levels data. In (4) we show results from Levinsohn-Petrin estimator, implemented using Stata levpet package

iii. Robust standard errors shown in parentheses

iv. *** p<0.01, ** p<0.05, * p<0.1

Table 1.7. Network Diversity, Stability and PageRank in the Hiring Network of IT and non-IT Labor

DV: Log(Sales)	(1)	(2)	(3)	(4)
Model	FE	FE	FE	FE
Log(Material)	0.572***	0.572***	0.587***	0.582***
	(0.0159)	(0.0159)	(0.0139)	(0.0144)
Log(Capital)	0.119***	0.118***	0.152***	0.139***
	(0.0149)	(0.0149)	(0.0125)	(0.0131)
Log(Non-IT Labor)	0.243***	0.243***	0.182***	0.212***
	(0.0185)	(0.0185)	(0.0119)	(0.0144)
Log(IT Labor)	0.0248***	0.0246***	0.0236***	0.0218***
	(0.00243)	(0.00241)	(0.00179)	(0.00183)
Log(IT Pool)	0.00839***	0.00841***	0.0138***	0.0132***
	(0.00309)	(0.00310)	(0.00358)	(0.00352)
Std(IT Net. Div.)	0.0139***	0.0131***	0.0102**	0.0112**
	(0.00499)	(0.00499)	(0.00439)	(0.00441)
Std(non-IT Net. Div.)		0.00457		-0.00464***
		(0.00489)		(0.00144)
Std(IT Net. Stability)	0.00229	0.00250		
	(0.00363)	(0.00366)		
Std(non-IT Net. Stability)		0.000389		
		(0.00444)		
IT PageRank			-0.00312**	0.00743
			(0.00148)	(0.00481)
Non-IT PageRank				0.00571***
				(0.00183)
Constant	0.264***	0.276***	0.472***	0.351***
	(0.0747)	(0.0757)	(0.0576)	(0.0647)
Controls	Industry Dummies, Year Dummies, Advertising Expense, R&D Expense, dummies for missing observations			
Observations	17,004	17,004	20,574	20,574
R-squared	0.9874	0.9874	0.9853	0.9856

Notes. i. Dependent variable in all regressions is log value of Sales. Regression also controls for year effect, advertising expenditure, R&D expense and dummy variables indicating where missing observations are filled in with industry average numbers
ii. Network Stability measures are calculated as the portion of hiring sources preserved from the previous year, calculated separately for IT and non-IT hiring
iii. Robust clustered standard errors shown in parentheses
iv. *** p<0.01, ** p<0.05, * p<0.1

Table 1.8. Industry Representation in the Labor Flow Network

DV: Log(Sales)	(1)	(2)	(3)
Model	FE	FE	FE
Log(Material)	0.558***	0.560***	0.557***
	(0.0161)	(0.0160)	(0.0160)
Log(Capital)	0.120***	0.121***	0.119***
	(0.0153)	(0.0153)	(0.0153)
Log(Non-IT employ)	0.253***	0.251***	0.251***
	(0.0188)	(0.0189)	(0.0188)
Log(IT employ)	0.0221***	0.0225***	0.0219***
	(0.00231)	(0.00232)	(0.00230)
Log(IT pool)	0.00943***	0.00922***	0.00927***
	(0.00294)	(0.00296)	(0.00295)
Std(IT Net. Div)	0.0157***		0.0145***
	(0.00487)		(0.00483)
#IT Industries represented	0.0242***		0.0219***
	(0.00414)		(0.00414)
Std(non-IT Net. Div.)		0.00409	0.00398
		(0.00459)	(0.00459)
# non-IT Industries represented		0.0253***	0.0175***
		(0.00641)	(0.00644)
Controls	Industry Dummies Year Dummies, Advertising Expense, R&D Expense, Dummies for missing observations	Industry Dummies Year Dummies, Advertising Expense, R&D Expense, Dummies for missing observations	Industry Dummies Year Dummies, Advertising Expense, R&D Expense, Dummies for missing observations
Constant	0.310***	0.295***	0.352***
	(0.0754)	(0.0802)	(0.0815)
Observations	16,707	16,707	16,707
R-squared	0.9863	0.9862	0.9863

Notes. i. Dependent variable in all regressions is log value of Sales. Regression also controls for year effect, advertising expenditure, R&D expense and dummy variables indicating where missing observations are filled in with industry average numbers
ii. #Industries represented by IT/ non-IT measures are constructed by counting the total number of different industries represented by the labor inflow
iii. Robust standard errors shown in parentheses
iv. *** p<0.01, ** p<0.05, * p<0.1

Table 1.9. Controlling for Quality of Employees

DV: Log(Sales)	(1)	(2)	(3)	(4)	(5)
Log(Material)	0.584***	0.584***	0.584***	0.584***	0.584***
	(0.0145)	(0.0146)	(0.0145)	(0.0145)	(0.0145)
Log(Capital)	0.140***	0.140***	0.140***	0.140***	0.140***
	(0.0133)	(0.0134)	(0.0133)	(0.0133)	(0.0133)
Log(Non-IT Labor)	0.210***	0.210***	0.210***	0.211***	0.211***
	(0.0150)	(0.0151)	(0.0150)	(0.0149)	(0.0150)
Log(IT Labor)	0.0236***	0.0237***	0.0233***	0.0238***	0.0234***
	(0.00200)	(0.00202)	(0.00203)	(0.00199)	(0.00203)
Log(IT pool)	0.0147***	0.0148***	0.0147***	0.0148***	0.0148***
	(0.00380)	(0.00381)	(0.00379)	(0.00379)	(0.00379)
Std(IT Net. Div)	0.00940**	0.00939**	0.00940**	0.00954**	0.00955**
	(0.00440)	(0.00441)	(0.00440)	(0.00443)	(0.00444)
Std(IT Stability)	0.00756	0.00754	0.00745	0.00773	0.00699
	(0.00497)	(0.00497)	(0.00499)	(0.00497)	(0.00497)
Avg. experience		0.000279			0.000964
		(0.00231)			(0.00249)
Avg. age			-0.00151		-0.00170
			(0.00112)		(0.00120)
University degree				0.0477	0.0496*
				(0.0293)	(0.0294)
Constant	0.346***	0.344***	0.392***	0.304***	0.351***
	(0.0668)	(0.0674)	(0.0738)	(0.0695)	(0.0763)
Controls	Industry Dummies, Year Dummies, Advertising Expense, R&D Expense, Dummies for missing observations				
Observations	20,344	20,344	20,344	20,344	20,344
R-squared	0.9857	0.9857	0.9857	0.9858	0.9858

Notes. i. Dependent variable in all regressions is the log value of Sales. Regression also controls for year effect, advertising expenditure, R&D expense and dummy variables indicating where missing observations are filled in with industry average numbers
ii. *Avg. experience* measures the average years of working experience for employees of the firm observed in our sample; *avg. age* is the average age of employees; *University degree* is the percentage of employees with holding a university degree
iii. Robust standard errors shown in parentheses
iv. *** p<0.01, ** p<0.05, * p<0.1

CHAPTER 2 : Data Analytics Skills and the Value of Social Media: an Organizational Complement Perspective

Abstract

Despite the rapid adoption of and increased spending on social media in the recent years, there is little existing research on the economic value of social media investment or the factors that affect this value. In this study, we first provide empirical evidence using a large sample of firms across industries to show firm market value increases with active social media usage, not just adoption. However, the return on using social media depends on having a larger number of employees with data analytics skills (rather than just IT skills), and that this complementarity is increased when employees with data analytics skills are dispersed throughout the firm. Overall, these results suggest that the value of social media for firms lies in its ability to facilitate the gathering and use of external data, and can be extracted more fully when a firm adjusts its human resource and organizational structure to facilitate data analysis and decision making.

2.1 Introduction

Social media has become an integral part of daily lives. As of 2014, more than 60% of Internet users have a Facebook account, and many other online consumers use specialized social media platforms such as LinkedIn, Twitter or Pinterest. User activities on Facebook are also growing rapidly. According to recent statistics, there are 4,166,677 Facebook "Likes" generated every single minute, almost doubling the amount in 2014.[9] The benefit of social media for companies' engagement with consumers could be tremendous. By interacting with customers and other stakeholders in social media, a firm can potentially reach new customers, increase engagement with existing customers, generate positive word of mouth for new and existing product offerings, and obtain timely and fine-grained data about customer preferences and behavior (Aral et al. 2013). However, the financial impact of social media still remains unclear. According to a 2014 marketing survey, most executives still do not know whether their social media expenditure has a positive return.[10] With corporate social media spending expected to take an increasing share of total marketing spending over the next years,[11] pressure is mounting to quantify the benefits of social media investment and identify the associated conditions and organizational factors needed so successfully leverage social media.

Prior work on information technology investments has stressed the importance of organizational complements, arguing that firms that are able to pair technology investments with appropriate investments in internal organization (e.g., decentralization) and human

---

[9] https://www.domo.com/blog/2015/08/data-never-sleeps-3-0/

[10] http://blogs.wsj.com/cmo/2014/09/03/social-media-spending-is-on-the-rise-but-impact-is-hard-to-measure/

[11] http://today.duke.edu/2014/09/cmosurveyaugust2014

capital receive greater value from these investments. It is likely that a similar relationship holds for social media investments, although the type of investments complementary to social media may differ from those shown to be complementary to other information technology-related investments. For instance, social media offers a unique opportunity to collect large-scale of real-time information about consumers, their activities and network relationships. In order to process such large amounts of new and detailed information, organizations need to have sufficient data processing and data analytics capabilities in place. We hypothesize that firms endowed with or investing in data-related capabilities are better able to incorporate social media into their firm strategy, and therefore are more able to capture value from these investments. We further argue that these capabilities are likely to be more important for social media technologies than for general IT or enterprise computing as examined in prior literature (Aral et al. 2012, Tambe et al. 2012). While many firms are experimenting with social media investments, not all have the required complementary data skills (or have the ability to rapidly acquire the requisite skills). This generates cross-sectional and time series variation in the apparent returns to social media investments, especially in forward-looking performance measures such as market valuation, which we will examine empirically in this study.

While existing studies on the impact of social media mostly focus on marketing outcomes (Chevalier and Mayzlin 2006, Dellarocas et al. 2007, Forman et al. 2008, Zhu and Zhang 2010), we argue that a company's ability to gather external information and its ability to process information internally are key to realizing the full benefits of social media strategies and need for these capabilities potentially extend to all operational areas of the firm. The information processing view of the firm (Galbraith 1977, Mendelson 2000, Radner 1992) suggests that a greater demand for organizational information processing created by the increased availability of external information can be met by either increasing the processing capability of individual decision makers or by delegating decision making more broadly. It follows from this theory that to meet the information processing needs generated by social media, firms can build capability by increasing the number of employees who have data analytics capability or expanding the range of departments that have access to and can act upon this newly available information. This argument leads to a prediction that social media investments and data skills are complements, and the strength of this complementary relationship is increased when data skills are more dispersed throughout the firm. While we believe we are the first to articulate the theory in this manner, industry observers have noted that the ability to extract value from social media may be closely tied to how data use is facilitated by organizational structure (Mims 2015).

In this paper we seek to complement and extend prior work on social media value in three specific ways. First, we examine the role that complementary organizational factors play in generating social media benefits and how these organizational factors are implemented and distributed throughout the firm can play a significant role in understanding how firms derive value from social media. Second, we utilize a market value framework, which reduces the need to make cost-related assumptions and is capable of making inferences on the long-term benefits of adopting social media. Lastly, we consider a broader sample—all publicly traded US firms—in contrast to focusing on only consumer-facing firms that dominate the existing literature. Our use of market value framework is similar to Luo et al. (2013) except that we rely on an econometric estimation rather than a prediction framework and to Chung et al. (2014) except we expand the framework to not only include social media outcomes but also organizational complements as the main mechanisms for social media's influence on firm performance. Because we consider all publicly traded firms, our sample is broader than comparable studies that focus on firms in consumer-facing industries, and thus we can provide a more comprehensive view on how social media use at enterprises affects productivity in broad sectors of the economy. In addition, the market value framework enables inferences about social media value beyond its direct impact on marketing outcomes. Overall, our approach allows us to examine the impact of social media on broad sectors of the economy, to examine whether the performance effects are heterogeneous across firms, and to determine if this heterogeneity is associated with organizational complements.

We compiled a 7-year panel (2007-2013) of social media adoption decisions by US publicly traded firms (10,171 firms in total). Using a combination of social media adoption and usage data collected from Facebook, publicly available financial information from Compustat, and measures of firm characteristics derived from a large database of employee information from online resumes (see e.g., Tambe (2014), Tambe and Hitt (2013), or Wu et al. (2014)), we are able to estimate the effect of social media activity on firms' market value, and the extent to which the presence of data analysis skills in the firm influences this value. We focus on Facebook because it is the most broadly used social network and the most likely place for a firm to make an initial social media investment (if only to establish a Facebook page).

Our empirical approach is based on estimating the market value effects of social media adoption using the Tobin's q framework, an econometric model that relates the market value of a firm to the quantities of the assets that the firm possesses (for examples of IT research applications of this framework see Bharadwaj et al. (1999) or Brynjolfsson et al. (2002)). We also control for firm, industry and time effects to limit unobserved heterogeneity, consistent with prior work utilizing this framework. This approach has several advantages for our study as it enables the detection of long-term value creation (in contrast to productivity analysis, which is better suited for identifying short-run effects) and is consistent with the view that analytic capabilities are assets whose value can be influenced by additional investment and firm strategy. The use of a market-value framework also enables the calculation of net benefits without having to make assumptions about the actual cost or investment in social media.

We find that firms that have data analytics capability, as measured by the skill sets of their employees, receive greater benefits from Facebook adoption. In fact, the data analytics skills outside of the marketing department are largely responsible for the positive return from using social media. To further explore this phenomenon, we find that the benefits are related to the dispersion of these capabilities across different departments of a firm. Moreover, the relationships we observe are unique to data analysis skills and not general IT skills, eliminating a potential confound and ruling out some types of endogeneity. Overall, these findings suggest that benefits of social media are higher in firms with a specific set of analytically-related capabilities, and that these capabilities are distinct from those that had been complementary to IT investments in the past.

The fact that social media adoption is complementary to some types of valuable capabilities but not others addresses some types some potential endogeneity issues, such as the possibility that "good" firms invest and adopt more leading edge technologies and skills. We also find that the direct effects of social media adoption are highest in industries with low levels of adoption, which is inconsistent with an alternative argument that our results are caused by firms in highly valued emerging industries (e.g., Internet companies) being early adopters of social media. Finally, further analysis of actual social media activities shows that the benefits only accrue from active social media usage, measured by either firm posting on Facebook page or consumer interactions, not simply the adoption event of starting a Facebook page. These observations increase our confidence that we are indeed observing marginal effects of social media use and its complements.

2.2 Literature Review

Our study connects three streams of research: (1) social media and firm performance, (2) the information processing theory and (3) organizational complementarities to technologies. Linking these theories, we show how firms with strong data analytics capabilities can derive greater value from using social media.

Measuring the Value of Social Media Activities for Organizations

Prior work on social media value, at least at the firm level, has focused primarily on linking social media use to various marketing outcomes (Chevalier and Mayzlin 2006, Dellarocas et al. 2007, Forman et al. 2008, Zhu and Zhang 2010). Studies have shown that social media can provide a variety of benefits including increasing brand recognition, facilitating consumer engagement, improving marketing strategy, and predicting product sales (Asur and Huberman 2010, Chen and Chellappa 2009, Ghose and Han 2011, Goh et al. 2013, Li and Wu 2014). Some have suggested that these benefits arise because marketing communications are more effective when transmitted through social network ties on social media platforms (Aral and Walker 2011, Bapna and Umyarov 2012). Others suggest that social media can more effectively promote word-of-mouth information diffusion not just among immediate social ties but also with a broader audience including friends of friends and beyond. This in turn can influence brand reputation and product sales (Chen et al. 2011, Dellarocas 2003, Li and Wu 2014). These studies typically focus on firms in consumer-facing industries where social media is likely to be used specifically for marketing purposes, and marketing metrics are especially meaningful and measurable.

Whereas customer satisfaction, brand recognition and product sales can help measure the effectiveness of social media on some marketing outcomes, they may not capture the full effects of social media on overall firm value. This is mainly because the cost of using social media is rarely observable; by observing only the benefits but not the costs, it is difficult to evaluate whether incremental sales of a social media campaign compare favorably with cost. While adopting social media is relatively inexpensive, maintaining an active social media presence and engaging consumers requires dedicated internal staff. This problem of estimating net value is further exacerbated when only the outcomes of a single product are observed (as it is in most social media studies), because it is hard to attribute social media cost to a single product when they can affect many different products and many of them could have complex interactions with each other (Desai 2001). Similarly, we cannot attribute costs to any longer-term influences on the brand or firm as a whole

(Agarwal et al. 2011). By adopting a market value approach in this study, we should be able to capture the long-run value of these investments (net of their costs) without having to identify or define the specific uses or outcomes for each individual firm. To check the validity of our theoretical and empirical approach, we stipulate our first hypothesis:

Hypothesis 1: *Social media adoption contributes positively to firms' market valuation.*

While there is a nacent branch of research that examines the connection between social media and overall firm value, the main intermediate mechanism examined is still marketing effectivness. Luo et al. (2013) studied whether social media activities can predict market value for product-oriented firms such as computer hardware and software firms. Chung et al. (2014) examined whether user-generated or firm-generated social media has a greater association with market value for 63 firms in South Korea that are also predominately consumer facing. While these studies are able to directly examine the effect of social media on firm performance, both studies were restricted to samples of firms in consumer-facing industries and focused on the effectiveness of marketing communications through social media.

We extend existing studies by implementing a market-value framework to firms across a wide range of industries. More importantly, we explore organizational complements and other important mechanisms that are beyond marketing strategy and effectiveness. This allows us to consider the overall impact of social media, including in the sectors where the impact is less understood, and get a more comprehensive view of the complements to social media.

Processing Information from Social Media and Data Analytics Skills

A recurring theme in the IT value literature is that the benefits of information technology investment disproportionately accrue to firms that have certain cluster of organizational complements (see a survey in (Galbraith 1977, Radner 1992). In particular, the information processing view of the firm (Galbraith 1977, Radner 1992) discusses how organizations can make more effective decisions when facing large amounts of available information. Earlier studies suggested two ways to deal with increased information load on an organization: improving the technical capacity for decision makers and/or distributing information throughout the organization (Mendelson 2000). This stream of research has identified the role of decentralized organizational structure, team-based incentives, and general human capital as complements to technology (Acemoglu et al. 2007, Bresnahan et al. 2002, Hitt and Brynjolfsson 1997, Tambe et al. 2012). Technology adoptions may also drive other complementary changes such as modified compensation systems (Aral et al. 2012). As certain types of complements become widely diffused and the capabilities of technology evolve, new organizational complements can become more important over time. For instance, recent work suggests that IT investment complements business practices that involve the gathering of information external to the firm (Tambe et al. 2012).

The rapid rise of corporate social media use potentially changes the value of different organizational complements just as prior generations of information technology-related innovations increased the value of general human capital and organizational decentralization. One unusual characteristic of social media is the very high volume of dynamic real-time data on individual interactions it can provide. For instance, a recent study (2014) suggested that Facebook users share 4.17 million pieces of content per minute[12] and the rate of social media information product continues to grow rapidly (Cisco 2014). In addition, social media operates on a shorter time scale (as fast as real time), enabling new types of decisions to be made that were not possible with product sales data or customer surveys. To effectively leverage social media, firms often need to continuously monitor and learn from the real-time data about their business. Thus, the ability to manage and process data is key to exploiting the full power of social media. Firms with stronger data analytics skills embodied in its workforce will be in better positions to process the information and effectively incorporate intelligence from social media data into their decision-making, and reap higher benefits from social media. By contrast, firms that do not have existing capabilities in handling data would be at a disadvantage in extracting the intelligence from real-time data. Presumably, firms can simply ramp up hiring once the need for data analytics is identified, but given the increased demand for analytics skills and other complementary changes firms would have to make to attract (and later retain) these types of employees, it is not obvious that these capabilities can be acquired quickly. Compared with traditional IT investments or enterprise software, social media investments can be initially quite small (although related marketing efforts and other external professional services can raise this substantially) and can be implemented on a very short time scale. With a much smaller barrier to entry, variations in returns to social media are

[12] https://web-assets.domo.com/blog/wp-content/uploads/2015/08/15_domo_data-never-sleeps-3_final1.png

much more likely to be driven by the ability to co-develop processes to collect, analyze and exploit the information these technologies provide. This leads to our third hypothesis:

Hypothesis 2: *Firms with capabilities relating to data management and analytics derive more value from social media.*

While social media data on consumer interactions could readily be used for marketing purposes, it also provides business intelligence in various aspects of a firm's operations, and has far-reaching impact on strategies beyond marketing effectiveness. For example, as users directly interact with firms' new product offerings through comments and reviews, firms can adapt their promotional strategies or even make product alterations (especially in information products industries).[13] Consumer sentiment on social media could be used to predict demand (Asur and Huberman 2010), which could then direct changes in firm's supply chain management. Sometimes, social media activities can also help firms to identify key talent to hire as many firms are starting to use their company page to turn some of their biggest fans into employees.[14] Benchmarking a firm's own social media activities against that of the competitors can help firms understand the general business environment and formulate strategies accordingly. Collectively, social media can help guide firms to flexibly adjust to changes in their business environment, hence, it is critical to tightly integrate social media data with firms' operation and overall strategy (Ghose et al. 2012).

Therefore, the ability of firms to capture value from social media, beyond simply acting as another business-to-consumer marketing channel, likely lies in a firm's ability to

[13] For instance, Zara built its strategy by embracing the fast-changing taste of consumers. By tapping on the pulse of consumer demands and preferences, Zara introduces new products twice every week (Petro 2012)
[14] https://www.smartrecruiters.com/blog/turn-your-facebook-fans-into-your-next-employees/

elicit, gather, process and act on social media-generated consumer information across different strategic areas of a firm's operation. This could suggest that a certain organizational decision-making structure is needed to coordinate decision-making processes across departments to better leverage social media. Just as decentralized decisions rights were shown to complement computing capabilities (Acemoglu et al. 2007, Bresnahan et al. 2002, Hitt and Brynjolfsson 1997, Tambe et al. 2012), the new capabilities of social media may be more effectively used by firms able to decentralize data analytics tasks across different departments to support decentralized decision making. When employees in various parts of the organization can understand and effectively incorporate intelligence from social media data into their decision-making, social media should have a larger effect on firm performance and achieving these benefits requires effort by employees outside the firms' marketing department. Consistent with the information processing view of the firm, we hypothesize that firms with data analytical skills are best able to leverage social media-generated information, and the capacity for use this information is further increased when these skills are distributed across the firm to enable decentralized decision making.

Hypothesis 3: *The value of social media is higher when data analysis skills are more dispersed throughout the firm.*

2.3 Empirical Framework

Data

We constructed our data sample from three primary sources: social media usage data derived from Facebook, the presence and distribution of various employees' skills (including analytic and IT skills) from a large scale database of employee resumes, and financial and related company information from Compustat. The final dataset consists of quarterly data for 10,171 companies across 7 years (2007-2013).

We use activities on Facebook as measure for social media engagement, since Facebook is among the earliest social media platform and is still one of the most popular social media sites. A firm page on Facebook is similar to an individual page. Firms "sign up" by providing their name, business category and location information after which they can post content, share information and interact with users. Users can comment, "like" or share information just as they do on individual sites.

We obtained from Compustat North America database a list of publicly traded firms still operating in 2007 (when Facebook started letting organizations and firms start their pages).[15] For each firm in the list, we queried Facebook to see whether it has a page, and if so, gathered information on the amount of firm- and user-generated content on the page. We used Facebook's search box to look up the corresponding page for each company name. If no relevant page showed up, we assume this particular firm has yet to start its Facebook page. If a page turns up in the search result, we performed an additional check to ensure that a correct match is found. For each Facebook page identified, we recorded the date when the page was created and analyzed all the content on the page to compute the total number of posts on the page, and the total number of likes, shares and comments from users in each quarter throughout our sample period. The number of posts indicates how

[15] A simple screening rules was is used to make sure all the observations are indeed from firms, instead of funds or public debts. Such rules include, for example, whether the firm reports the value of its fixed assets, or the number of employees.

frequently the firm is using the social media platform to publish new information. User interactions with company posts, as reflected by clicking to "like" or "share" on a post, or issuing comments, would capture the level of individual users' engagement with the firm's social media activities. These metrics of social media engagement has been shown to affect performance and marketing outcomes (Chung et al. 2014). From these data we calculated two measures to use in our analysis. ***Facebook Adoption*** is a 0/1 measure taking the value of 1 after the firm has created a Facebook page. ***Facebook Engagement*** is an index variable created from the first principal component of the number of likes, comment and shares on the Facebook page, since these 3 measures are highly correlated with one another.

Among our sample of 10,171 firms, a total of 1,921 (around 19%) are found to have Facebook pages over the sample time period. In Table 2.1, we summarize the total number and the percentage of Facebook adopters in each industry, as well as the industry breakdown for all the Facebook adopters. The industry classification is performed using Standard Industrial Classification (SIC) codes at the "1.5 digit" level to identify 13 different industries (see Table 2.1), all of which are well represented in our sample. Overall, we find that retailers and computer-related firms are generally more likely to have Facebook pages, consistent with expectations.

In order to test for how data skills among employees influence firms' ability to obtain value from social media usage, we use a database consisting of more than 6 million individual resumes in 2007. These data are similar to other large sample resume datasets used for prior work in IT value and technology diffusion (see e.g., (Tambe and Hitt 2013, Tambe and Hitt 2013)) for a more detailed discussion of the advantages and limitations of these datasets generally). Using professional tools designed for parsing resumes, we identified all the skills indicated in each resume and when these skills were obtained. We then aggregated the skills of all employees at a particular firm and quarter (and also at the department level in each firm) conservatively assuming that the individual obtained all skills shown in her resume by the time of the most recent employment. We define ***Data Analytics Skills*** as "data centric analytics skills" and "data mining skills" identified by the resume parser.[16] The job titles of individuals with data analytics skills are broad and span multiple business areas. They include consultant, financial analyst, systems engineer, customer service specialist, program manager, and systems analyst among others. We categorized these job titles into several departments such as sales, engineering, finance, administrative, research & development or manufacturing, to calculate the distribution of data skills across different sectors of the firm. An individual defined to have ***IT Skills*** if her most recent employment is an IT-related job. Specifically, we identify employees with IT skills as either having a job title clearly associated with information technology (e.g. software engineer, systems analyst, programmer analyst) or indicating relevant keywords related to IT elsewhere on their resume (e.g., computer, software, web development). Both IT skills and data skills measures are normalized by the total number of employees at the firm. Various alternative measures of identifying IT or job skills were also explored and they did not qualitatively change our results.[17] Using similar methods, we also account for

---

[16] One tool we used was the Sovren resume parser (see www.sovren.com). This tool is used by a number of online job sites to facilitate skills-based resume searches (among other activities).

[17] Using the same method to identify IT skills as we did with data skills would be too fuzzy, for employees could easily have some level of basic programming or office software skills and thus confound the measure.

employees with communication, marketing and consumer relationship skills for each firm. We match this cross-section of aggregated firm level skills information in 2007 to all firms in our panel across all years.

Finally, we link the data on social media usage and employee skills to quarterly financial data from Compustat from 2007 to 2013. Since our data suggest that the majority of the firms that joined Facebook did so between 2009 and 2012, the time span of our panel should be adequate for identifying relationship between social media use and subsequent changes in market value. The primary dependent variable in this analysis is firm market value, which is calculated as the sum of the market value of equity (based on stock prices at the end of the period) plus the book value of debt. The primary control variables are fixed assets (property plant and equipment), other assets (principally financial assets and intangibles), R&D assets, and advertising assets. The asset values of R&D and advertising are constructed respectively using R&D and advertising expense in each period through a perpetual inventory method that has been employed in prior research (Hall 1990, Hirschey and Weygandt 1985). Our primary models also include industry and time controls derived from these data. These measures are similar to those used in prior studies of IT value based on a similar framework (Brynjolfsson et al. 2002, Brynjolfsson and Yang 1999, Hitt and Brynjolfsson 1996). Summary statistics for the variables and their correlations are reported in Tables 2.2 and 2.3.

Methods

Under the "q theory" of investment (Tobin 1969), a firm should invest in assets until the marginal value of an additional dollar of the asset is equal to a dollar of market value.

This ratio of market value of a firm to total book value is known as Tobin's q. While theory implies that it is the marginal value of q that should be approximately one, it is commonly assumed in empirical work that the average value of Tobin's q is a good approximation for the marginal value (Hayashi 1982). This implies an estimating equation of the form:

$$Market\ Value = \sum_{i \in [Asset\ Types]} \alpha_i A_i$$

Where A_i represents the quantity (book value or investment cost) of different assets and α_i is the marginal value (which should be 1 in equilibrium for each asset). Essentially, this equation suggests that the value of firm is the sum of the value of its assets. To implement this equation empirically, the framework is to relate market value to the book value of fixed assets, other assets, R&D and advertising assets (Brynjolfsson et al. 2002). In this study, we first incorporate Facebook Adoption and Facebook Engagement measures into the equation metrics. Our basic regression is:

$$\begin{aligned} Market\ Value = {} & \alpha + \beta_1 Fixed\ Assets + \beta_2 Other\ Assets + \beta_3 AdvertisingAssets \\ & + \beta_4 R\&DAssets + \ \gamma_1 FacebookAdoption + industry\ controls \\ & + time\ controls + \ \varepsilon \quad (1) \end{aligned}$$

We measure the effect of Facebook adoption on firms' market value through the coefficient γ_1. In the next step, we explore how data skills in an organization influences the social media strategies for firms, by adding the Data Skills and IT skills measures and their interaction with Facebook Engagement into the regression.

We use the Least Absolute Deviations (LAD) regression for most of our analysis, consistent with prior work on estimating market value models in heterogeneous data

(Brynjolfsson et al. 2002). The linear in levels relationship between assets and market value combined with substantial cross-firm variation in size can cause OLS to perform poorly in market value regressions, especially when large firms have characteristics that deviate significantly from population mean. By weighting the residuals by the absolute value rather than the square, this method is less sensitive to outliers. However, there are few panel data variants of LAD and none appear to perform well for panels with a large cross-sectional dimension, so our use of LAD will tend to underestimated standard errors due to repeated sampling of the same firm over time. In the worst case, where every firm is exactly the same in every time period, this will tend to lower the standard error by the square root of the time dimension (although this is typically less in practice), so we will generally consider only effects that are strongly statistically significant.

2.4 Results

Social Media Adoption and Firm Valuation

In Table 2.4, we present the baseline estimates of our market value regression. As described in equation (1), on top of the standard market value regression framework relating market value to four basic asset measures (fixed assets, other assets, advertising assets and R&D assets), we add a binary variable that takes the value of 1 following the creation of a Facebook public page for that firm ("*Facebook Adoption*"). We find that the market value of a firm is $4 million higher following Facebook adoption (column 1). The value of fixed assets is approximately 1 as implied by theory and we find other assets to be "worth" approximately $0.5 to $0.7 per dollar of book value which is similar to prior estimates of this equation in other work (Brynjolfsson et al. 2002). In column 2, we estimate the same model but restricting the sample to only eventual Facebook users. Although the sample is reduced substantially (omitting about 80% of the data for firms that do not have a Facebook page), results are similar to column 1.[18] Thus we use the smaller sample to examine how different measures of firm engagement (as represented by number of posts) and user engagement (as in the number of likes, shares and comments) can affect market value (column 3-6). Since all usage measures are demeaned with a standard deviation of one, the result in column 3 suggest that a one standard deviation increase in posting frequency is associated with a $10 million in market value when compared to firms with an average number of posts. Interestingly, as soon as we introduce measures of social media use in the model, the direct effect of Facebook adoption becomes negative and the Facebook engagement metrics are positive and statistically significant. The fact that the direct effect of Facebook adoption turns negative helps rule out some types of reverse causality. If the positive return on adopting social media is driven by highly valued firms adopting the technology early, this would imply a positive direct effect on Facebook adoption.

Different metrics of Facebook engagement show different thresholds for when Facebook adoption value turns positive – for the lesser forms of engagement (posts) the firm does not achieve positive value until approximately 0.6 standard deviations above the mean (column 3), while firms that have more than 0.1 standard deviations above the mean

[18] Since our Facebook Adoption measure takes the value 1 after firm creates a Facebook page, we further exam whether this is mainly driven by firms that eventually adopt Facebook vs. firms that never use Facebook page. Using a “difference in difference” style estimate (adopters vs. non adopters; and post adoption vs. non-adoption for the eventual adopters). This returns results similar to what is shown in columns (1) and (2) of Table 2.4.

in likes are earning positive returns to adoption (column 4). While these results are suggestive of benefits of user engagement, we remain cautious about their interpretation since the user interactions may also be capturing latent factors such as how engaged users are, which could be heterogeneous across firms. The correlations among the three user engagement variables are 0.55 or greater, suggesting they may be measuring the same underlying mechanism. To simplify interpretation, we construct a composite measure of user engagement from the first principal component of these three measures (the composite explains 68% of the variance). Estimates for the resulting variable ("*Facebook Engagement*") suggest that firms with one standard deviation higher user engagement have about a $48 million higher market value (column 7).

While we have found that the market value effect mainly comes from social media usage and not just adoption, it is still possible that some types of highly valued firms are naturally a better fit for social media and therefore they may simultaneously have higher social media activities and higher market value. To examine this effect, we divide firms into two groups by their industry: those that are involved in the production or sales of consumer products and services and therefore more likely to use social media to attract and interact with their consumers (referred to as "*Consumer-Related Group*"); the remainder, which consists of other industries not directly dealing with the end consumers (the "*Non-Consumer Related Group*"). The consumer-related group is made up mainly of three types of industries: retail, computing and consumer products manufacturing. We identified consumer related group by computing an index of Facebook usage and Internet display advertising usage, and selecting industries that had the highest composite. By contrast, the non-consumer related group places less priority on consumer interactions since they do not directly deal with end consumers. This classification is also corroborated by the actual Facebook adoption rates in the two groups. Among the consumer related group, 38.6% have adopted a Facebook page, in contrast to 16.4% in the low non-consumer related group. The χ^2 test show that the between group difference is statistically significant ($p < 0.0001$) suggesting that on average the adoption rates for the two groups are different. If the effect of social media use on market value is being driven by industries with a large Facebook presence we would expect a stronger effect for the consumer-related rather than the non-consumer-related group.

In Table 2.5, we show that for firms in the consumer-related group, the marginal benefit of social media adoption on their market value is not different from zero. In contrast, for firms in the non-consumer related group, having a Facebook page slightly improves market value. On average, Facebook adoption is associated with adding $4 million in market value (column 2). When we compare only within the group of eventual Facebook users it appears that the effect is driven by engagement rather than just adoption. Interestingly, we also find that firms in the non-consumer related group benefit more from social media engagement than firms in the consumer related group (columns 5-6). These results are not consistent with a simple selection story where firms that are inherently more likely to benefit from social media on average choose to adopt Facebook. Rather, our results indicate the opposite—that there are marginal benefits for starting a social media campaign only in industries where the social media presence is not already the norm. This is perhaps not surprising—having a Facebook page may be a competitive necessity in consumer-facing industries, while a Facebook page might provide a (small) source of differentiation in industries where customer engagement through online media is less the norm. Facebook could be providing a new channel for these firms to engage their customers who would not normally engage with a firm online. Overall, these results support Hypothesis 1.

Social Media and Data Analytics Skills

While having a Facebook company page is shown to increase firms' market valuation on average, the effect across firms is likely to vary. We hypothesized that the value of social media use would be associated with a firm's data analytics capabilities as embodied in the skills of its workforce. To test for this mechanism, we examine the how the observed value of social media engagement varies with a firm's data analytics capabilities.

Table 2.6, column 1 verifies the previous results on social media adoption (see Table 2.4, column 7) still holds for the subsample of firms with both financial information and skills information available. In column 2, we examine the performance effect of having data analytics skills within the company's workforce and its interaction with social media engagement. Data analytics skills not only contribute directly to higher firm value, generating $49 million for each standard deviation increase in data analytics skills, but also magnify the effect of Facebook engagement. With each standard deviation increase in data analytics skills, a firms sees an additional $89 million value increase when firms simultaneously engage in social media activities on Facebook. This confirms our Hypothesis 3 that data analytics skills complements social media usage.

An alternative hypothesis is that data skills are simply a proxy for overall IT investment. To rule out this possibility, in columns 3 and 4 of Table 2.6, we examine whether general IT skills would capture all the effect and complementarities observed with data skills. We measure IT skills as the portion of IT workers among all the employees included in our sample for each period. Column 3 shows that existing IT talent in the work force contributes positively to firm valuation and also positively interacts to some degree with Facebook engagement, but column 4 shows that data skills and its interaction with social media dominates the effect of IT skills. It is clear that data skills are distinct from general IT skills, playing a unique role in increasing firm value from social media strategies. Columns 5-8 show similar results on the subsample of eventual Facebook users.

While the quantile regression framework performs well for this type of market value model, it does not easily provide a mean to control for firm level effects (especially given the large number of firms we consider). To verify that our results are not entirely driven by firm level differences, we also estimate a linear fixed effect model (see Table 2.7). These results further confirm that after controlling for firm-level fixed effects, we still observe that data analytics skills complement social media usage, while general IT skills do not (see columns 1-4). The results are consistent using the subsample of eventual Facebook users as shown in column 5-8. In summary, we have found that data analytics skills are unique (at least compared to general IT investment) in generating complementarities with social media. This new complementarity is distinct from earlier complements that are found to be associated with general information technology investment.

Social Media and Analytic Skills beyond the Marketing Department

We have shown so far that in our data social media presence and active user engagement increase market value and especially for firms with existing data capabilities. In the next step, we seek to verify whether social media influences firms' performance through more than just the marketing channel. While the marketing related areas are the more obvious use of social media data, other aspects of a firm's operations, like product design, human resource management, demand forecasting, and supply chain management could also benefit from data collected from social media. To achieve these strategic goals, non-marketing departments also need to process social media data effectively, and we expect the data analysis abilities in these departments also positively interact with firms' social media usage.

We identify employees in our sample as from marketing or non-marketing department according to their job titles. Typical marketing job titles include keywords or phrases like "marketing", "advertising", "brand manager", and "promotion supervisor." We then calculate the percentage of employees with data analytics skills in the marketing department and the non-marketing departments respectively. In Table 2.8, we replicate columns on data skills from Table 2.6 for easy comparability. Column 3 shows that data skills in both marketing and non-marketing departments positively influence the market valuation for the firms; each standard deviation contributes to around $30 million increase in market value for both types. Columns 4-6 show that data skills in different departments interact with social media use differently. Specifically, having data skills in non-marketing departments can substantially benefits from social media use, with each standard deviation increase in social media usage adding an extra $103 million in market value. Data skills in marketing departments, on the other hand, do not seem to complement social media usage. Column (7) includes additional controls for IT skills and its interaction with social media use. Again we observe social media use to complement only with data skills in non-marketing departments but not with data skills in marketing department nor with firm's overall IT skills.

These results collectively suggest that social media is not simply acting through the marketing channel in creating value. Data collected from social media data can help guiding strategies in other departments and yield far-reaching impact across different areas of business. Therefore, firms with data analytics skills across different departments have higher information processing capacity and stand better chances of reaping benefits from social media through all the potential channels.

Social Media Usage and Dispersion of Data analytics Skills

We have found so far that active usage of social media improves an organization's market valuation; this is consistent with an argument that data analytics skills play a significant role in transforming information collected from social media to value-creating enterprise strategies and such a role extends beyond the marketing channels. Since social media is one of the main channels through which firms interact with users, it is likely that decentralized data analytics skills across the organization could be important for each department in leveraging the data collected from social media to more effectively engage in the decentralized decision making.

To examine this possibility, we categorize employees in the organization into 7 departments according to their job titles: 1) manufacturing, 2) engineering, 3) sales and marketing, 4) human resources, 5) finance and accounting, 6) research and development, and 7) administrative. Then, we measure the percentage of data skills attributed into each of the departments and construct the Herfindahl–Hirschman Index (HHI) to quantify the dispersion of data skills in the organization. Higher value of HHI represents a more concentrated data skill distribution while a lower score suggests that analytics skills are dispersed throughout the firm. Given the way the measure is defined, if HHI has a positive relationship with firm performance, a centralized data analytics structure is more beneficial. However, if HHI is found to be negatively correlated with firm performance, a dispersion of data skills is more beneficial.

$$HHI = \sum_{i \in departements} {s_i}^2$$

$$s_i = \frac{\#employee\ with\ data\ skills\ in\ department\ i}{\text{total}\ \#\ employees\ with\ data\ skills}$$

We explore the effect of dispersion of data skills in the organization and its interaction with social media usage and report the results in Table 2.9. Column 1 shows that more dispersed data analytical skills across departments (lower HHI) increase firm value. Furthermore, column 2 shows that dispersion of data skills positively interacts with social media engagement ($\beta_{\text{HHI * Facebook}} = -15.37$, $\rho<0.01$). This is consistent with our earlier results that firms with employees equipped with data analytical skills across various departments are more effective in using the information from social media. Column 3 confirms that this effect is in addition to the complementarities between the overall data skills in the firm and social media usage. Again, similar results are observed for the subsample of eventual Facebook users (Columns 4-6). Results are also similar with other common dispersion metrics such as entropy (not shown).

These results support Hypothesis 4 that a decentralized distribution of data analytics skills across different departments is important for effectively processing the information and enabling firm to derive value from social media in many potential channels. This finding is consistent with the information processing theory that decentralized decision making helps meet the needs of increased inflow of data. It is also supported by some anecdotal evidence that start-ups are gradually shifting towards decentralized management systems, relying on the increasing amount of data available to each individual department to support this decentralization (Mims 2015). Because firms with decentralized data analytical skills across the organization are more capable of processing the social media data for decision making in many different business aspects, they stand to benefit more from social media usage.

2.5 Robustness Checks

Both Facebook adoption and the presence of data analysis skills are potentially subject to various forms of reverse causality. Of some concern is the standard "free cash flow" argument that suggests that firms with more slack resources will have higher market value and also be more willing to invest in innovative practices. Our results suggest this type of confound is unlikely because the effect on market value from social media come from engagements and not mere adoption. In addition, our primary results rely on complementarities arguments–while it is easy to believe that both adoption and data skill acquisition are endogeneous, it is more difficult to argue how this would predict that greater value is accrued when they are used together. We also show firms in industries that are less likely to use social media experience greater benefits from the technology. This finding further suggest that the role of this particular form endogeneity is limited. This also reduces the likelihood of our results driven by industry-related selection effects.

The fact that the results do not hold for IT skills also reduces the likelihood of reverse causality or omitted variables biases. If data analytics skills were simply a proxy for firms' skills related to general technology and social media were just a proxy for general technology investment, the complementarities we observe could just be evidence for the general complementarities between IT investment and IT skills (Tambe and Hitt 2013). Instead we find that general IT skills do not complement social media while data analytics skills do. Collectively, these results suggest that the complementaries between data analytics skills and social media use is unique and distinct from earlier complements associated with IT. Only when firms possess both social media and have substantial capabilities in handling the data can they maximize their return on using social media.

Additional unobserved endogeneity concerns could also be present in our setting. For example, a firm's social media adoption and usage may be endogeneous because there could be unobserved characteristics driving both the firm's market valuation and its ability to generate a significant social media presence. We use four sets of instruments to address this potential bias. First, we construct an instrumental variable as the residual of regressing Facebook engagements that were incurred at the time of adoption on advertising expense. This residual represents the propensity to adopt Facebook free of marketing effects and thus can be viewed as a proxy for the latent demand to adopt Facebook. We use Facebook engagement measured during a narrow window around the time of the adoption to capture the inherent interest of users to engage the firm on social media without needing to address the effect from subsequent promotional activities that may also increase user engagements. Second, we use the average social media usage of other firms from which the focal firm hires workers and to which current employees are hired away. These measures reflect users' latent willingness to engage with certain kinds of product or certain types of businesses. Similar firms likely have similar social media engagements, but these hiring neighbors should not directly affect the market value of the focal firm. Third, we use the average Facebook engagement of the surrounding firms located in the five counties closest to the focal firm's headquarter—the firm decision to use Facebook is likely influenced by their geographic neighbors' decisions, but its market valuation should not be influenced by regional neighbors' social media activities. Finally, we construct an instrument representing the general online presence of other firms with which the focal firms engage in hiring activities. We measure online presence based on the historical traffic rank on Alexa (provided by Amazon web services) of a firm's home page. The online presence of the hiring neighbors is correlated with the focal firm's social media engagement, because they may have similar characteristics that drive similar user engagement. However, neighbors' social media engagements are unlikely to directly affect the market value of the focal firm.

Employee skills in different firms could also be endogeneous; high market value may signal that the firm is an attractive place to work, especially for workers with desired skills.

Since the data we have on employee skills is cross-sectional, we treat skills as quasi-fixed with respect to market valuation. Relative to market valuation or IT investments, the composition of employee skills within a firm, especially large public traded firms in our sample, is difficult to change in a short amount of time. Therefore, we consider skills to be quasi-fixed with respect to technology adoption and performance (Applegate et al. 1988, Bresnahan et al. 1999, Brynjolfsson and Hitt 1996, Milgrom and Roberts 1990).

We repeat the analysis, instrumenting for social media usage and its interactions with data skills and IT skills. Since there is no effective framework for instrumental variables for using quantile regressions in large datasets, we use panel data IV regressions with fixed effects and random effects as well as a dynamic GMM model with external instrumental variables. The fixed effects panel IV results (column 4, Table 2.10) and the random effect panel IV results (column 6) show the direction of the estimates is consistent with their respective baseline regression results (columns 3 and 5). These results suggest that data analytics skills in employees improve firm valuation and also amplify the benefits of social media activities. Furthermore, this result is not driven by more successful firms being better at engaging users on social media or other unobserved characteristics about the firm that can increase both social engagement and market value. We are reluctant to interpret the effect size in these models here because the panel setup is not the best to capture the relationship between market value changes and the social media strategies[19]. This can be further exacerbated after controlling for firm level effects because the residuals may not

[19] The estimates on fixed assets and other assets are not close to the theoretical value while the LAD estimates do.

contain much information to effectively capture the effect of Facebook engagement on market valuation.

Next, we use the Arellano-Bond/Blundell-Bover two-step robust system GMM estimation with all available lags for the endogenous variables and we also included external instrumental variables to address other forms of reverse causality between market valuation and Facebook engagement. This procedure uses appropriate internal panel instruments lagged levels and differences) to estimate differences and levels regressions and then optimally weights them using the sample error matrix estimated from the first-step regression, which gives efficient estimates that are robust to firm heterogeneity. The Arellano-Bond text for AR(2) in the first differences, the Hansen J statistics (over-identification test) and the difference in difference Hansen test verify the validity of the instruments used in the GMM estimation. In the dynamic GMM estimation, we continue to observe that data analytics skills amplify the effect of Facebook engagement on market valuations (column 7).

In addition, we used similar dynamic GMM method with external IVs to check the result on data skills dispersion and social media engagement. Results are reported in columns (2) and (4) in Table 2.11. Here, we are treating the Herfindahl–Hirschman Index of data skills dispersion across departments in the firm as quasi-fixed and instrument for Facebook engagement. We continue to observe strong support for the hypothesis that data skills dispersed across the organization facilitates information processing and enable the firms to benefit more from social media activities.

2.6 Conclusion

In this study, we examine the relationship between social media use, complementary organizational and industry factors, and firm market value. This work builds on prior work on social media value by expanding the pool of firms, conducting the analysis in a market value framework, which does not require cost estimates to calculate value and can capture long-run benefits (at least those perceived by outside investors), and developing measures for plausible complementary investments or capabilities. We are among the first studies to examine the impact of social media beyond the marketing channel and to explore the role of information processing capacity, as represented by the data analytics capabilities of the workforce, in reaping benefits from social media.

Overall, our baseline results suggest that social media investments are valuable in general, at least to the extent that adoption is followed by actual consumer or firm use. Moreover, our measured marginal benefits of social media appear higher in consumer-facing industries, suggesting that firms in industries not directly engaging with consumers might also start experimenting with social media strategies. Furthermore, we find support for our core hypothesis–that analytics skills are complementary to social media usage. While we find that data skills and social media use are associated with higher value generally, firms that combine social media adoption with data skills receive an additional benefit. Such benefit is unique to data analytics skills, because general IT skills do not appear to be strongly complementary to social media usage. These findings suggest that social media requires different organizational complements than general IT investments. In order to meet the new needs of processing large amounts of real-time data collected from social media, firms should work on improving their information processing capability, for example, starting with having more employees with data analytical skills for its talent pool.

In addition, we examine the distribution of data skills across various departments of the organization for more insight into how social media creates firm value. Our results suggest that data skills outside the marketing area of the firm are a stronger complement to social media use, consistent with the argument that it is the ability to utilize social media-generated information for firm operations that is driving firm value, rather than using social media as a novel or lower cost marketing channel. Furthermore, we find a more dispersed distribution of data skills across departments in the entire firm to positively affect firm productivity as well as a complementarity between widespread analytics capability and social media use. This view, which is consistent with prior theories on organizational decentralization and information processing in organizations, further reinforces prior work that suggests the ability to process and utilize external information is an important complement to modern information-technology innovations. While most of the focus so far has been on social media for marketing related purposes, in fact, firms should also look for potential uses for information collected from social media across different areas of business. Preparing different departments to effectively analyze social media data could help firms receive higher benefits from social media usage beyond the marketing channel.

Overall, from a research perspective our results suggest an important distinction between organizational complements to social media adoption and usage and those found in prior work on IT adoption, which can inform future research on the value of social media. Moreover, we also show an important role of organizational complementarities in explaining the cross-sectional returns to social media investment, and the potential benefits of examining social media investments outside of consumer products industries where much of the existing research has been focused. From a managerial standpoint, our results suggest that firms can generally gain greater value from social media with appropriate complementary investments in decentralized use of data analytics skills, and that a significant benefit of social media lies in the ability to use social media information to support firm decisions.

2.7 References

Acemoglu D, Aghion P, Lelarge C, Van Reenen J, Zilibotti F (2007). Technology, Information, and the Decentralization of the Firm. Quarterly Journal of Economics. 122(4) 1759-1799.

Agarwal A, Hosanagar K, Smith MD (2011). Location, location, location: An analysis of profitability of position in online advertising markets. Journal of Marketing Research. 48(6) 1057-1073.

Applegate LM, Cash JI, Mills DQ (1988). Information Technology and tomorrows manager. Harvard Business Review. 66(6) 128-136.

Aral S, Brynjolfsson E, Wu L (2012). Three-way complementarities: Performance pay, human resource analytics, and information technology. Management Science. 58(5) 913-931.

Aral S, Dellarocas C, Godes D (2013). Introduction to the special issue-social media and business transformation: A framework for research. Information Systems Research. 24(1) 3-13.

Aral S, Walker D (2011). Creating social contagion through viral product design: A randomized trial of peer influence in networks. Management science. 57(9) 1623-1639.

Asur S, Huberman B (2010). Predicting the future with social media. IEEE.

Bapna R, Umyarov A (2012). Do your online friends make you pay? A randomized field experiment in an online music social network. NBER working paper.

Bharadwaj AS, Bharadwaj SG, Konsynski BR (1999). Information technology effects on firm performance as measured by Tobin's q. Management science. 45(7) 1008-1024.

Bresnahan T, Brynjolfsson E, Lorin MH (2002). Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. Quarterly Journal of Economics. 117 339-375.

Bresnahan TF, Brynjolfsson E, Hitt LM (1999). Information technology, workplace organization and the demand for skilled labor: Firm-level evidence. National Bureau of Economic Research.

Brynjolfsson E, Hitt L (1996). Paradox lost? Firm-level evidence on the returns to information systems spending. Management science. 42(4) 541-558.

Brynjolfsson E, Hitt LM, Yang S (2002). Intangible assets: Computers and organizational capital. Brookings papers on economic activity. 2002(1) 137-198.

Brynjolfsson E, Yang S (1999). The intangible costs and benefits of computer investments: Evidence from the financial markets. Citeseer.

Chen C, Chellappa R (2009). Music sales and online user activity at social networking sites. Citeseer.

Chen Y, Wang Q, Xie J (2011). Online social interactions: A natural experiment on word of mouth versus observational learning. Journal of Marketing Research. 48(2) 238-254.

Chevalier JA, Mayzlin D (2006). The effect of word of mouth on sales: Online book reviews. Journal of marketing research. 43(3) 345-354.

Chung S, Animesh A, Han K, Pinsonneault A (2014). Firms' Social Media Efforts, Consumer Behavior, and Firm Performance: Evidence from Facebook. Consumer Behavior, and Firm Performance: Evidence from Facebook (June 10, 2014).

Dellarocas C (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. Management science. 49(10) 1407-1424.

Dellarocas C, Zhang XM, Awad NF (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. Journal of Interactive marketing. 21(4) 23-45.

Desai PS (2001). Quality segmentation in spatial markets: When does cannibalization affect product line design? Marketing Science. 20(3) 265-283.

Forman C, Ghose A, Wiesenfeld B (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. Information Systems Research. 19(3) 291-313.

Galbraith JR (1977). Organization design: An information processing view. Organizational Effectiveness Center and School. 21 21-26.

Ghose A, Han SP (2011). An empirical analysis of user content generation and usage behavior on the mobile Internet. Management Science. 57(9) 1671-1691.

Ghose A, Ipeirotis PG, Li B (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. Marketing Science. 31(3) 493-520.

Goh K-Y, Heng C-S, Lin Z (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. Information Systems Research. 24(1) 88-107.

Hall BH (1990). The manufacturing sector master file: 1959-1987. National Bureau of Economic Research.

Hayashi F (1982). Tobin's marginal q and average q: A neoclassical interpretation. Econometrica: Journal of the Econometric Society 213-224.

Hirschey M, Weygandt JJ (1985). Amortization policy for advertising and research and development expenditures. Journal of Accounting Research 326-335.

Hitt LM, Brynjolfsson E (1996). Productivity, business profitability, and consumer surplus: three different measures of information technology value. MIS quarterly 121-142.

Hitt LM, Brynjolfsson E (1997). Information technology and internal firm organization: An exploratory analysis. Journal of Management Information Systems 81-101.

Li X, Wu L (2014). Herding and Social Media Word-of-Mouth: Evidence from Groupon. Available at SSRN 2264411.

Luo X, Zhang J, Duan W (2013). Social media and firm equity value. Information Systems Research. 24(1) 146-163.

Mendelson H (2000). Organizational architecture and success in the information technology industry. Management science. 46(4) 513-529.

Milgrom P, Roberts J (1990). The economics of modern manufacturing: Technology, strategy, and organization. The American Economic Review 511-528.

Mims C (2015). Data is the New Middle Manager. Online.

Radner R (1992). Hierarchy: The economics of managing. Journal of economic literature 1382-1415.

Tambe P (2014). Big Data Investment, Skills, and Firm Value. Management Science. 60(6) 1452-1469.

Tambe P, Hitt LM (2013). Job hopping, information technology spillovers, and productivity growth. Management Science. 60(2) 338-355.

Tambe P, Hitt LM (2013). Measuring information technology spillovers. Information Systems Research. 25(1) 53-71.

Tambe P, Hitt LM, Brynjolfsson E (2012). The extroverted firm: How external information practices affect innovation and productivity. Management Science. 58(5) 843-859.

Tobin J (1969). A general equilibrium approach to monetary theory. Journal of money, credit and banking. 1(1) 15-29.

Wu L, Jin F, Hitt LM (2014). Are All Spillovers Created Equal? A Network Perspective on IT Labor Movements.

Zhu F, Zhang X (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. Journal of Marketing. 74(2) 133-148.

Table 2.1. Facebook Adoption by Industry

Industry	# Facebook Users	# Total Firms	Portion of FB user in Industry	Portion of All FB Users
Durable Manufacturing	362	1,839	19.68%	18.84%
Mining	267	1,702	15.69%	13.90%
Finance, Insurance & Real Estate	224	1,681	13.33%	11.66%
Non-durable Manufacturing	254	1,492	17.02%	13.22%
Computer, Software	222	904	24.56%	11.56%
Services, except Financial	157	691	22.72%	8.17%
Utilities	146	684	21.35%	7.60%
Retail Trade	149	420	35.48%	7.76%
Transportation	61	238	25.63%	3.18%
Wholesale Trade	35	228	15.35%	1.82%
Public Administration	21	182	11.54%	1.09%
Construction	15	79	18.99%	0.78%
Agriculture, Forestry & Fishing	8	31	25.81%	0.42%
Total	1,921	10,171	18.89%	100.00%

Table 2.2. Summary Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Market Value	115,715	4,752.40	27,352.92	0	971,689
Fixed Assets	115,715	2,390.91	13,814.15	0	493,970
Other Assets	115,715	2,487.59	34,272.07	0	2,000,478
Facebook Adoption	115,715	0.07	0.25	0	1
Number of Posts (Per Quarter)	115,715	0.77	4.36	0	105
Total Likes(Per Quarter)	22,627	209.67	1,108.93	0	26,718
Total Comments(Per Quarter)	22,627	33.91	220.76	0	7,281
Total Shares(Per Quarter)	22,627	24.72	214.79	0	13,748
Total Employees	94,662	91.97	422.77	0	8,049
IT Skills	94,662	0.06	0.13	0	1
Data Skills	94,662	0.21	0.24	0	1
Data Skills in Marketing Department	94,662	0.01	0.06	0	1
Data Skills in Non-marketing Departments	94,662	0.19	0.24	0	1
HHI for Data Skills	94,662	784.29	2204.73	0	10,000

Table 2.3. Correlations between Main Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.Market Value	1.000													
2.Fixed Assets	0.573	1.000												
3.Other Assets	0.631	0.066	1.000											
4.Facebook Adoption	0.022	0.003	0.008	1.000										
5.Number of Posts	0.013	0.012	0.013	0.694	1.000									
6.Total Likes	0.019	0.030	0.001	0.367	0.491	1.000								
7.Total Comments	0.014	0.018	0.012	0.304	0.391	0.653	1.000							
8.Total Shares	0.009	0.010	0.001	0.209	0.272	0.582	0.298	1.000						
9. Total Employees	0.325	0.266	0.129	0.036	0.023	0.071	0.073	0.028	1.000					
10. IT Skills	0.004	-0.022	0.004	0.014	0.025	-0.017	-0.009	-0.010	-0.014	1.000				
11. Data Skills	0.053	0.027	0.036	0.002	0.002	-0.019	-0.010	-0.011	0.007	0.107	1.000			
12. Data Skills in Marketing Department	-0.008	-0.016	-0.002	0.030	0.022	-0.002	0.000	0.001	0.010	0.001	0.229	1.000		
13. Data Skills in non-marketing departments	0.062	0.036	0.041	-0.005	-0.003	-0.019	-0.010	-0.011	0.008	0.115	0.947	-0.037	1.000	
14.HHI for Data Skills	-0.053	-0.052	-0.026	-0.003	-0.003	-0.015	-0.003	-0.014	-0.053	0.035	0.146	0.143	0.118	1.000

Note: Measures of skills are normalized by the total number of employees at the firm;

Table 2.4. Facebook Usage and Market Value of Firms

DV: Market Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample	All	FB_users	FB_users	FB_users	FB_users	FB_users	FB_users
Fixed Assets	0.968***	0.935***	0.924***	0.924***	0.924***	0.925***	0.924***
	(2.38e-05)	(9.85e-05)	(8.09e-05)	(8.13e-05)	(6.94e-05)	(7.08e-05)	(7.91e-05)
Other Assets	0.662***	0.356***	0.606***	0.607***	0.608***	0.607***	0.607***
	(9.07e-06)	(3.61e-05)	(5.10e-05)	(5.12e-05)	(4.38e-05)	(4.46e-05)	(4.98e-05)
Facebook Adoption	4.265***	7.650**	-6.510*	-6.553*	-5.838**	-4.854	-6.754**
	(1.255)	(3.661)	(3.470)	(3.485)	(2.977)	(3.033)	(3.388)
Number of Posts			10.00***	3.942***	6.095***	7.005***	3.836***
			(1.393)	(1.492)	(1.235)	(1.241)	(1.455)
Total Likes				52.90***			
				(1.221)			
Total Comments					36.55***		
					(0.996)		
Total Shares						35.60***	
						(0.989)	
Facebook Engagement							48.60***
							(1.195)
Other Controls	Advertising Asset, R&D Asset, Industry, Year and Quarter, Dummies for Missing Values						
Constant	17.43***	144.4***	150.7***	162.4***	146.4***	143.4***	161.7***
	(4.510)	(16.33)	(14.56)	(14.62)	(11.48)	(11.69)	(14.21)
Observations	115,715	22,627	22,627	22,627	22,627	22,627	22,627

Notes: i. Column 1 uses all observations; columns 2-7 use only firms that have a Facebook page
ii. Number of Posts, Total Likes, Total Comments and Total Shares measures are standardized
iii. Facebook Engagement is a principle component of Total Likes, Total Comments and Total Shares
iv. Advertising Assets, R&D Assets are calculated as cumulative expense over the years, using 0.136 and 0.149 respectively as discounting rate; these numbers are from Hirchey & Weygandt (1985)
v. Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 2.5. Facebook Usage in Consumer Related/non-Consumer Related Industries

DV: Market Value	(1)	(2)	(3)	(4)	(5)	(6)
Sample	All	All	FB_users	FB_users	FB_users	FB_users
Sub Group	Consumer Related	Non-Consumer Related	Consumer Related	Non-Consumer Related	Consumer Related	Non-Consumer Related
Fixed Assets	1.242***	0.950***	1.176***	0.893***	1.176***	0.891***
	(0.000344)	(1.76e-05)	(0.000438)	(5.56e-05)	(0.000486)	(5.36e-05)
Other Assets	1.721***	0.625***	2.094***	0.351***	2.094***	0.351***
	(0.000265)	(6.63e-06)	(0.000560)	(2.18e-05)	(0.000621)	(2.10e-05)
Facebook Adoption	3.846	4.152***	-8.498*	1.511	-8.446	-1.837
	(4.631)	(1.035)	(4.599)	(2.293)	(5.282)	(2.234)
Facebook Engagement					-0.725	42.06***
					(1.940)	(0.853)
Other Controls	Advertising Asset, R&D Asset, Industry, Year and Quarter, Dummies for Missing Values					
Constant	-110.6***	59.77***	5.783	160.4***	5.808	149.9***
	(5.674)	(3.518)	(12.85)	(10.70)	(13.63)	(9.839)
Observations	16,805	98,910	4,627	18,000	4,627	18,000

Notes: i. The consumer related group includes firms in the following industries: 1) retail 2) computer related 3) consumer product related manufacturing; the non-consumer related group includes the rest

ii. Columns 1-2 use all sample; columns 3-6 use the sample firms with a Facebook page

iii. Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 2.6. Skills and Technology Implementation

VARIABLES	(1) All	(2) All	(3) All	(4) All	(5) FB users	(6) FB users	(7) FB users	(8) FB users
Fixed Assets	0.952***	0.952***	0.953***	0.952***	1.002***	1.000***	1.001***	0.999***
	(0.000265)	(0.000260)	(0.000225)	(0.000261)	(0.000868)	(0.000655)	(0.000722)	(0.000703)
Other Assets	0.351***	0.351***	0.351***	0.351***	0.458***	0.455***	0.460***	0.457***
	(0.000137)	(0.000134)	(0.000116)	(0.000135)	(0.000278)	(0.000210)	(0.000231)	(0.000225)
Facebook Adoption	-58.70***	-69.38***	-67.63***	-73.10***	-45.81*	-58.49***	-49.82**	-68.15***
	(17.35)	(17.03)	(14.77)	(17.15)	(25.28)	(19.02)	(21.03)	(20.44)
Facebook Engagement	61.06***	88.73***	72.65***	87.92***	88.22***	144.8***	87.24***	145.1***
	(4.540)	(4.737)	(4.366)	(5.128)	(9.374)	(7.636)	(8.786)	(8.603)
Data Skills		48.82***		48.35***		73.41***		76.26***
		(4.192)		(4.229)		(6.805)		(7.340)
Facebook Engagement * Data Skills		88.82***		89.54***		125.7***		157.6***
		(6.803)		(7.260)		(11.18)		(12.80)
IT Skills			12.23***	4.749			17.27**	5.341
			(3.726)	(4.350)			(7.670)	(7.507)
Facebook Engagement * IT Skills			11.95*	-4.650			-1.275	-55.71***
			(6.991)	(8.628)			(14.64)	(14.80)
Other Controls	Advertising Asset, R&D Asset, Industry, Quarter, Dummy of Missing Values							
Constant	601.9***	500.5***	503.1***	493.5***	558.8***	503.7***	472.6***	486.9***
	(29.09)	(19.56)	(17.05)	(19.91)	(60.16)	(30.84)	(34.37)	(33.46)
Observations	35,909	35,909	35,909	35,909	8,662	8,662	8,662	8,662

Notes: i. Data and IT skills are respectively measured by the percentage of employees with the skill
ii. Data skills are identified as data-centric software skills and data mining skills
iii. Columns 1-4 use the whole sample; columns 5-8 use the sample firms with a Facebook page
iv. Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 2.7. Skills and Technology Implementation (Fixed Effects Model)

DV: Market Value	(1) All	(2) All	(3) All	(4) All	(5) FB_users	(6) FB_users	(7) FB_users	(8) FB_users
Other Assets	0.527***	0.412***	0.412***	0.412***	0.422***	0.474***	0.476***	0.474***
	(0.0495)	(0.0360)	(0.0360)	(0.0360)	(0.0999)	(0.0663)	(0.0664)	(0.0663)
Facebook Adoption	1,543***	199.3	246.2	199.8	67.10	42.34	58.08	48.54
	(339.4)	(191.9)	(191.5)	(192.0)	(173.2)	(170.8)	(170.7)	(170.6)
Facebook Engagement		163.5**	42.07	164.1**		253.9**	123.1	228.2**
		(68.11)	(46.33)	(67.51)		(116.3)	(98.09)	(110.9)
Facebook Engagement		333.2***		334.5***		634.6***		663.4***
* Data Skills		(98.54)		(98.43)		(183.9)		(209.7)
Facebook Engagement			-16.24	8.196			146.4	-108.5
* IT Skills			(26.03)	(29.74)			(151.0)	(172.3)
Other Controls	Advertising Asset, R&D Asset, Industry, Quarter, Dummy of Missing Values							
Constant	1,840***	2,717***	2,720***	2,716***	2,295**	2,249**	2,245**	2,218**
	(687.3)	(502.4)	(502.1)	(502.3)	(1,103)	(1,031)	(1,042)	(1,035)
Observations	32,543	32,437	32,437	32,437	8,673	8,662	8,662	8,662
R-squared	0.952	0.927	0.927	0.927	0.946	0.937	0.937	0.937

Notes: i. Dependent variable is market value minus fixed assets, i.e. setting the coefficient for fixed assets at the theoretical value of 1

ii. Data and IT skills are respectively measured by the percentage of employees with the skill

iii. Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 2.8. Data Skills in Marketing and Non-Marketing Departments

DV: Market Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fixed Assets	0.952***	0.952***	0.951***	0.952***	0.952***	0.952***	0.952***
	(0.000265)	(0.000265)	(0.000255)	(0.000264)	(0.000260)	(0.000259)	(0.000256)
Other Assets	0.351***	0.351***	0.351***	0.351***	0.351***	0.351***	0.351***
	(0.000137)	(0.000137)	(0.000131)	(0.000136)	(0.000134)	(0.000133)	(0.000132)
Facebook Adoption	-58.70***	-69.02***	-66.04***	-62.66***	-71.67***	-74.20***	-81.12***
	(17.35)	(17.36)	(16.69)	(17.26)	(17.05)	(16.94)	(16.85)
Facebook Engagement	61.06***	90.52***	69.60***	65.62***	97.93***	100.3***	99.24***
	(4.540)	(4.962)	(4.363)	(4.530)	(4.745)	(4.737)	(5.069)
Data Skills		48.68***					
		(4.218)					
Facebook Engagement * Data Skills		90.63***					
		(7.080)					
Data Skills in Marketing Department			29.53***	30.62***		26.43***	26.72***
			(4.045)	(4.265)		(4.176)	(4.142)
Facebook Adoption * Data Skills in Marketing Dept				12.19		-31.82***	-31.76***
				(12.56)		(12.31)	(12.22)
Data Skills in non-Marketing Department			35.08***		46.89***	45.03***	44.14***
			(4.095)		(4.200)	(4.170)	(4.160)
Facebook Adoption * Data Skills in non-Marketing Dept					103.1***	111.0***	112.2***
					(6.820)	(6.787)	(7.172)
IT Skills							6.177
							(4.275)
Facebook Engagement * IT Skills							-4.734
							(8.503)
Other Controls	Advertising Asset, R&D Asset, Industry, Year and Quarter, Dummy of Missing Values						
Constant	601.9***	498.4***	493.1***	586.2***	508.4***	497.8***	490.0***
	(29.09)	(19.94)	(19.16)	(28.95)	(19.58)	(19.43)	(19.54)
Observations	35,909	35,909	35,909	35,909	35,909	35,909	35,909

Notes: i. The division of data skills into marketing and non-marketing departments is based on individuals' job title, for example, job titles containing keywords like "sales, marketing, advertising, brand manager" are considered marketing job titles;
ii. Clustered standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 2.9. Dispersion of Data Skills and Social Media Usage

DV: Market Value	(1) All	(2) All	(3) All	(4) FB_users	(5) FB_users	(6) FB_users
Fixed Assets	0.958***	0.959***	0.959***	1.003***	1.003***	1.003***
	(0.000355)	(0.000364)	(0.000345)	(0.00104)	(0.000980)	(0.000989)
Other Assets	0.351***	0.351***	0.351***	0.443***	0.444***	0.443***
	(0.000183)	(0.000187)	(0.000178)	(0.000334)	(0.000313)	(0.000316)
Facebook Engagement	80.20***	80.11***	105.3***	122.1***	138.3***	184.5***
	(6.151)	(6.339)	(6.410)	(11.28)	(10.60)	(11.25)
Data Skills	50.87***	52.69***	61.10***	44.00***	45.03***	80.20***
	(5.815)	(5.967)	(5.695)	(10.62)	(9.956)	(10.31)
HHI for Data Skills	-26.74***	-27.42***	-28.53***	-29.46***	-36.73***	-40.11***
	(5.754)	(5.907)	(5.615)	(10.59)	(9.991)	(10.03)
HHI* Facebook Engagement		-15.47**	-30.54***		-28.95**	-54.45***
		(7.165)	(7.692)		(12.96)	(12.34)
Data Skills * Facebook Engagement			101.3***			176.5***
			(9.411)			(17.32)
Other Controls	Advertising Asset, R&D Asset, Industry, Quarter, Dummy for Missing Values					
Constant	249.4***	206.5***	210.1***	386.4***	352.9***	356.1***
	(38.87)	(26.91)	(25.55)	(72.10)	(45.59)	(46.07)
Observations	35,064	35,064	35,064	8,551	8,551	8,551

Notes: i. HHI is calculated as data skills distributed across the following departments: 1) manufacturing 2) engineering 3) sales and marketing 4) human resource 5) accounting and finance 6) R&D 7) administrative;
ii. Clustered standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 2.10. Robustness Check

DV: Market Value	(1) OLS	(2) Quantile	(3) FE	(4) FE/IV	(5) RE	(6) RE/IV	(7) GMM/IV
Fixed Assets	0.879***	0.952***	0.525***	0.526***	0.622***	0.617***	0.792***
	(0.00415)	(0.000261)	(0.0734)	(0.00870)	(0.0722)	(0.00757)	(0.0498)
Other Assets	0.300***	0.351***	0.404***	0.404***	0.357***	0.358***	0.750***
	(0.00233)	(0.000135)	(0.0353)	(0.00537)	(0.0467)	(0.00408)	(0.0358)
Facebook Engagement	125.5*	87.92***	122.8**	480.9***	135.4*	480.8***	331.2***
	(71.95)	(5.128)	(50.94)	(117.3)	(77.06)	(118.8)	(40.61)
Data Skills	410.1***	48.35***			437.0**	442.7**	138.8
	(63.42)	(4.229)			(198.8)	(218.7)	(135.7)
Facebook Engagement * Data Skills	726.1***	89.54***	189.0***	648.0**	225.0**	694.3**	93.70**
	(110.4)	(7.260)	(68.19)	(287.0)	(110.1)	(289.6)	(38.95)
IT Skills	103.7	4.749			-116.3	256.1	123.6
	(64.50)	(4.350)			(71.90)	(220.3)	(120.2)
Facebook Engagement * IT Skills	103.4	-4.650	-90.16**	-180.6	-99.03*	-341.3**	-220.4***
	(107.0)	(8.628)	(40.24)	(162.5)	(57.22)	(169.3)	(27.96)
Other Controls	Advertising Asset, R&D Asset, Industry, Quarter, Dummy for Missing Values						
Constant	11,111***	493.5***	3,499***	3,687***	3,208*	2,920***	17,610***
	(663.6)	(19.91)	(1,056)	(621.6)	(1,783)	(859.0)	(2,652)
Observations	35,769	35,909	35,769	35,769	35,769	35,769	35,769

Notes: i. IVs for Facebook Engagement include: 1) Facebook Engagement in the first quarter since the firm adopts Facebook page; 2) Average social media engagement in other companies from where the focal firm hires from and employees move to (i.e. network neighbors); 3) Average social media engagement in other companies located in the 5 counties closes to the focal firms' headquarter; 4) Historical website traffic to homepage of network neighbor firms; their second order terms and all cross terms.

ii. Column 1 reports results from OLS regression, column 2 replicates the previous results from quantile regression in Table 6 to here; column 3 reports results from panel regression with fixed effects; column 4 uses dynamic panel instrument variable regression with fixed effects and instrumenting for Facebook engagement and its interactions with data skills and IT skills; bold font indicates variables instrumented for; similarly, column 5-6 use the random effects framework; column 7 reports results using two-step robust system GMM, with the internal panel instruments and external IVs we specified;

iii. Clustered standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 2.11. Robustness Check for Dispersion of Data Skills and Social Media Usage

DV: Market Value	(1) All Quantile	(2) All GMM/IV	(3) FB_users Quantile	(4) FB_users GMM/IV
Fixed Assets	0.959***	0.851***	1.003***	1.269***
	(0.000345)	(0.0467)	(0.000989)	(0.0773)
Other Assets	0.351***	0.748***	0.443***	0.649***
	(0.000178)	(0.0298)	(0.000316)	(0.0343)
Facebook Engagement	105.3***	344.2***	184.5***	937.0***
	(6.410)	(46.67)	(11.25)	(110.5)
Data Skills	61.10***	83.94	80.20***	-110.3
	(5.695)	(125.1)	(10.31)	(128.7)
HHI for Data Skills	-28.53***	-75.65	-40.11***	73.67
	(5.615)	(70.27)	(10.03)	(87.25)
HHI* Facebook	-30.54***	-110.4***	-54.45***	-130.7***
Engagement	(7.692)	(22.45)	(12.34)	(22.70)
Data Skills * Facebook	101.3***	-2.170	176.5***	304.2**
Engagement	(9.411)	(40.25)	(17.32)	(122.3)
Other Controls	Advertising Asset, R&D Asset, Industry, Quarter, Dummy for Missing Values			
Constant	210.1***	210.1***	356.1***	356.1***
	(25.55)	(25.55)	(46.07)	(46.07)
Observations	35,064	35,064	8,551	8,551

Notes: i. HHI is calculated as data skills distributed across the following departments: 1) manufacturing, 2) engineering, 3) sales and marketing, 4) human resource, 5) accounting and finance, 6) R&D, and 7) administrative;

ii. Columns 1 and 3 show the previous quantile regression results from Table 2.9; columns 2 and 4 show results two-step robust system GMM, with the internal panel instruments and external IVs we specified. Facebook Engagement and its interaction terms with HHI and with Data Skills are treated as endogenous and instrumented for.

iii. Clustered standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

CHAPTER 3 : Social is the New Financial: How Startups' Social Media Activities Influence Funding Outcomes

Abstract

Early state firms are increasingly utilizing social media to communicate with customers and potential investors. We investigate whether the use of social media is associated with increased success in raising venture capital. Social media can potentially improve startup funding success in two ways: 1) enabling investor discovery of potential investment opportunities and 2) providing additional information to investors that enables a better evaluation of the quality of the venture. Using social media activities on Twitter and venture financing data from CrunchBase, we find that active social media presence and strong Twitter influence (followers, mentions, impressions, and sentiment) increase startups funding success, amount raised and breadth of investor pool. We also find that social media has a greater association with raising capital from investors with less access to information (e.g., angels) and who are less industry specialized, consistent with an improvement in investors ability to discover potential investments. In addition, we find that the relationships are stronger for startups where quality information may be less available such as firms outside geographic venture capital clusters or where later investors do not have network relationships with early investors, consistent with an additional information channel to evaluate startup quality.

3.1 Introduction

With 72% of U.S. Internet users on Facebook and 23% on Twitter, social media has become an important conduit of information for individuals, firms, and markets. Social media provides an alternative channel for marketing communication between firms and customers, enabling firms to build their brands and interact with customers. The effectiveness of social media for marketing goods and services has been particularly well-studied in the context of established firms (Aral et al. 2013, Bharadwaj et al. 2013). However, few studies have looked at the use of social media by emerging firms or the role of social media in the capital markets. In this project, we explore the intersection of those two areas to study whether social media activities improve a startup's ability to raise capital from venture capitalists and angel investors, arguably one of the most important factors in the success of early stage firms.

The market for early stage private financing faces two distinct information challenges, both of which social media can address. First, startup firms seeking private equity or debt financing are not listed in centralized exchanges like publically traded firms, and investors need to engage in costly search for to identify potential startups to finance. Startups that are "off the radar" due to their location or lack of existing relationships between officers and potential investors are therefore less likely to receive funding without alternative means of communication. Second, startups lack traditional assets and cash flows histories, which presents a challenge for investors to evaluate their quality (Aldrich and Fiol 1994). This is exacerbated by a principal-agent conflict where entrepreneurs are incentivized to exaggerate potential growth and earnings to attract investors (Dessein 2005). Historically, these information deficiencies have been overcome by geographic agglomeration of investors and startups (See, e.g. Saxenian 1991), allowing for more informal contact between investors and startup officers, and information networks among investors who may have participated together in deals in the past (Hochberg et al. 2007). Here social media provides an additional source of information that does not rely on geography or existing social networks among investors.

Social media acts as a medium for information exchange and can offer solutions to both the costly search and lack of information problems that market participants face. Startup firms can broadcast information about themselves over social media and thus raise awareness of their existence among potential investors, helping investors discover more early stage ventures and expand their consideration set of potential investment opportunities. In addition, startups' social media activities can provide investors with an additional information channel for evaluating investment opportunities. For example, popularity on social media could demonstrate a startup's ability to attract specific customer groups, build its brand name, and integrate feedback from consumers. Such positive social media information would signal firm quality to investors and raise their expected return on the investment, and thus increase the startup's chances of receiving larger amounts of funding. Anecdotally, venture capitalists are increasingly conducting "due diligence" on social media platforms and reacting favorably to startups with effective social media performance. For example, Vandaele Capital LLC decided to fund Boxtera, a startup which delivers health-food packages to subscribers, because of their effective use of Twitter to reach their target audience.[20]

Regulators are also taking note of social media's growing role as a conduit for investment information. Historically, startups were restricted in their ability to make public offers or solicitations to sell shares or securities, including on social media platforms. However, the substitution of social media for traditional information sources, such as press releases, has introduced ambiguity into the definition of appropriate communications to potential investors. Along with the implementation of various other provisions of the J.O.B.S. Act,[21] the U.S. Securities and Exchange Commission (SEC) issued a new policy in June 2015 allowing startups to tweet about their investment opportunities to potential investors, something they were not previously allowed to do on public platforms. As this channel of communicating with potential investors gains further legitimacy, it is more important that we understand the policy implications for early stage venture financing markets and for entrepreneurial performance.

Existing studies on social media mostly focuses on the marketing outcomes and on established firms. Studies have shown that social media can promote word-of-mouth information diffusion (Aral et al. 2013, Chevalier and Mayzlin 2006, Dellarocas et al. 2007,

[20] Wall Street Journal. "If You Look Good on Twitter, VCs May Take Notice". September 30, 2013. http://www.wsj.com/articles/SB10001424127887324659404578499702279196058

[21] Another provision of the Jumpstart Our Business Startups (JOBS) Act is the legalization of equity crowdfunding, the online offering of private equity securities to investors. While various forms of crowdfunding are likely to occupy a growing persistent component of the market for early stage equity financing, traditional venture capital and angel financing are expected to continue to dominate early stage private equity financing market for the foreseeable future. Nevertheless, our research on the implication of social media for venture capital and angel financing should also have implications for the future of social media in equity crowdfunding as well.

Forman et al. 2008, Zhu and Zhang 2010) and serve as a platform for greater consumer engagement with a product or brand (Chen and Chellappa 2009, Chen et al. 2015, Ghose and Han 2011, Goh et al. 2013, Li and Wu 2014). Recent studies further link social media activity and firm performance through mechanisms of marketing effectiveness (Chung et al. 2014, Luo et al. 2013) and value extraction from social media analytics (Hitt, Jin and Wu, 2015). However, there is limited work directly examining the use of social media by and its effect on early stage firms, with the notable exception of related work by Aggarwal et al. (2012), who examine social media mentions of a firm (particularly blogs) and venture financing; they find that the impact of negative electronic word-of-mouth is greater than is the impact of positive word-of-mouth, and that the effect on financing decreases as a firm progresses through the stages of financing.

This paper bridges the information systems literature and the entrepreneurial finance literature, providing empirical evidence for the effect of startup firms' social media activities on their funding outcomes. We construct a unique data set that combines financing rounds data for high-technology startups, as reported in CrunchBase, with historical data on Twitter activity by the same startups, from Topsy.com. We empirically investigate distinct hypotheses from the two mechanisms by which social media may facilitate entrepreneurial financing, specifically, how social media helps investors discover potential investment opportunities though search cost reduction, and how social media activity provides an additional channel of information for investors to assess startup quality.

We find that in general, social media activity on Twitter improves startups' chances of successfully getting funded from investors, raises the total amount of funding they receive, and increases the number of investors investing in the firm, after controlling for various firm-level characteristics. In addition, we find evidence supporting both of the two channels through which social media can influence startup funding. Startups active on social media are likely to attract a larger portion of angel investors in early funding rounds. Since angel investors are usually not full-time investors, and have less alternative channels for obtaining information about possible investment opportunities as compared to VCs, social media plays a larger role in their discovery of potential startups to finance. Also, we find that startups active on social media attracts a larger portion of investors with diverse investment portfolios rather than concentrated investments in specific industries. Since investors making repeated investments in the same industries generally build up connections and sources of information to learn about new investment opportunities, social media's role as an information broadcasting channel is less significant. On the other hand, for investors making diverse investments, social media provides a low cost mechanism to become aware of startup activities across a range of industries.

We also find evidence that startups' social media activities give investors information to better evaluate potential investment opportunities. Specifically, we find that startups located outside the VC cluster regions (Boston, New York, San Francisco) see more significant increase in the size of funding from social media activities. We argue that while investors' decision on whether to invest or not is likely influenced by discovery of the startups, the funding size reflects on investors' valuation of the startup firm. Therefore, increased funding from social media activity to startups located in regions where it is harder for investors to inspect the startups on site, shows that positive information on social media can reduce uncertainty in startup quality and improve investors' valuation for the startups. In addition, we find that startups active on social media are more likely to receive funding from experienced investors, especially when they do not have trusted sources of quality information through investor syndicate networks. Since experienced investors are likely to have accumulated expertise analyzing startup quality from prior investments, they should be more effective analyzing social media information to assess startups. Especially when investors do not have partners from previous joint investments already invested in the startups as information source, social media plays a more important role as an additional information channel that enables better evaluation of startups' quality.

This study provides insight into how social media serves as an alternative channel of information and network connections for startups and their investors, and this work has several practitioner implications: entrepreneurs should take advantage of the new SEC regulations and effectively leverage social media campaigns to seek investors. In addition to broadcasting their presence to potential investors, entrepreneurs should also manage their social media content to portray a positive brand image, demonstrate the ability to engage target customer segments, and source informative customer feedback. From the investors' perspective, social media presents them with an alternative channel to discover startup firms, particularly if they do not have connections or channels of information in certain industries. Therefore, investors should consider the quality of startups as revealed by their social media activities, such as their interactions with consumers, in making their investment decisions.

3.2 Theoretical Background and Hypotheses

Private equity investments by venture capital firms and angel investors continue to be a dominant source of financing of early stage, high-growth, high-risk, technology businesses. In 2014, annual venture capital inflows topped $48 billion for these "startups", representing the highest levels in over a decade.[22] The two most common types of investors are venture capital firms that invest into startups using funds put up by institutional or large private investors, and "angel investors" who are high-net-worth individuals[23] investing their own funds.[24] These financial intermediaries specialize in the evaluation, investment execution, and operation of startups.

Venture capital firms and angel investors face unique information challenges in discovering and evaluating investment opportunities. First, startups lack many of the traditional physical assets and steady cash flow histories used to evaluate more established businesses (Aldrich and Fiol 1994), so investors have less information and substantial uncertainty when evaluating a new venture (Kaplan and Strömberg 2004, Shane and Cable 2002). The information problem facing investors is further exacerbated by an asymmetric

[22] PricewaterhouseCoopers LLP and National Venture Capital Association. January 16, 2015. http://nvca.org/pressreleases/annual-venture-capital-investment-tops-48-billion-2014-reaching-highest-level-decade-according-moneytree-report/

[23] In the United States, angel investors must be accredited by the SEC, meaning they must have a net worth of at least $1 million (not including the value of their primary residence) or have an income over $200k each year for the last two years.

[24] New forms of entrepreneurial finance, such as peer-to-peer lending and crowdfunding, have developed in the last few years, but they continue to be a niche segment of the capital market for early stage private equity in both scale and influence.

information problem between entrepreneurs and investors (Dessein 2005), that entrepreneurs have an incentive to over-represent the quality of their firm to investors in the hope of improving their chance at receiving an investment at a high valuation. Thus, hard information for the evaluation of new ventures is rare and the marginal value of additional information is likely to be high (Amit et al. 1990, Gompers 1995).

Second, the lack of a centralized market for early stage private equity means that entrepreneurs and investors lack information about the existence of parties on the other side of the market, and thus they must undergo a costly search process (Inderst and Müller 2004) in order to identify a possible choice set before they can even begin the process of information collection and evaluation (due diligence). In other words, possible investors may not even be aware of a particular venture and ventures may have limited knowledge of available funding opportunities. These search costs can be prohibitive, making legitimate high-quality ventures unable to locate funding or do so on acceptable terms.

Social media could alleviate both of these information problems by broadcasting information about the existence startup seeking financing to potential investors and by offering another channel of information for investors to evaluate startup quality through their social media activities.

Impact of Social Media on Organizations

While the literature on social media use by emerging firms is still nascent, many of the same insights from this literature apply to social media usage of early stage firms.

Perhaps the most studied aspect of social media is its role as a new marketing channel to customers. A substantial literature has linked feedback from consumers, such as product reviews, to product sales and changes in marketing strategy (Dellarocas et al. 2007, Forman et al. 2008, Hong et al. 2014, Li and Hitt 2008, Zhu and Zhang 2010). In addition, social media contributes to long-run marketing performance, by providing an alternative channel for organizations to build brand names and encourage consumer engagement (Ghose and Han 2011, Goh et al. 2013, Lee et al. 2014, Rishika et al. 2013, Shriver et al. 2013). Finally, social media can provide a platform for new marketing strategies such as encouraging the diffusion of information and social influence through network ties (Angst et al. 2010, Aral and Walker 2011, Bapna and Umyarov 2015). In addition to documenting the direct performance of social media on marketing outcomes, a new stream of research demonstrates that these benefits translate into improvements in overall firm performance (Chung et al. 2014, Hitt et al. 2015, Luo et al. 2013).

An emerging stream of studies shows the impact of social media on the online financing market, mostly looking at the crowdfunding or peer-to-peer lending markets (Agrawal et al. 2011, Lin et al. 2013). Eesley and Wu (2015) study the diversity of entrepreneurs' social network connection with mentors, the motivation for these relationships, and their influence on startup firm performance. Greenwood and Gopal (2015) examine how media coverage of different technology segments influences the number of new startups founding in that sector. Aggarwal et al. (2012) find a link between blog mentions and sentiment about startups and subsequent financing outcomes.

Overall, existing studies suggest that the use of social media can influence overall firm performance, including the success of early stage firms and their ability to obtain financing. Therefore, we hypothesize that startups more active on social media are more likely to succeed in their funding process:

Hypothesis 1: *Startups active on social media are more likely to receive* ***larger amounts*** *of funding from investors.*

Hypothesis 2: *Startups active on social media are more likely to receive funding from* ***a larger number*** *of investors.*

The next two sections outline the specific mechanisms, namely search cost reduction and startup quality information channel, that together serve to drive Hypothesis 1 and Hypothesis 2.

Social Media, Search costs and Discovery of Investment Opportunities

Unlike publicly traded companies, early stage startup firms do not have a centralized market where investors have easy access to all potential investment opportunities. In fact, private equity investors and entrepreneurs engage in a costly search process to find one another (Inderst and Müller 2004). Furthermore, there exist few brokers connecting startups with investors, such as investment bankers for more mature companies. Search cost remains a significant issue preventing investors from discovering the full set of investable startups.

A long stream of information systems literature examines the role of IT in reducing search costs. Digital communications technologies can substitute for geographic proximity, enabling firms to locate "closer" to customers or their target markets without incurring significant costs of coordination or uncertainty (Bakos and Brynjolfsson 1993, Clemons and Row 1992, Gurbaxani and Whang 1991, Malone et al. 1987). Like prior advances in information technology, social media presents a similar opportunity for low cost communication, particularly for early stage firms that are less likely to have developed the more traditional marketing channels that require more upfront capital investment.

Twitter, in particular, enables firms to broadcast information to a targeted audience (Chen et al. 2015). Twitter's open platform design enables users to follow anyone they wish, making it an effective channel to distribute information to a large group of already interested users (Fischer and Reuber 2011). Just as established firms can use Twitter to reach to customers, startup firms can also use it to broadcast about themselves and reach out to a larger pool of both customers and investors. Since investors have easy access to information on Twitter, this low cost information channel can broaden their pool of potential investment opportunities.

While social media information is equally available to all types of investors, it may be especially valuable to investors lacking the usual information channels to know about potential investment opportunities. Angel investors, for example, do not typically dedicate a substantial amount of time to sourcing possible investments and do not have teams of research staffs or access to institutional information that are usually available to venture capital firms (Lin et al. 2015). Since angel investors' usually face higher search cost finding out potential startups to invest in, we expect to see social media playing a larger role in their discovery of new investment opportunities.

Hypothesis 3: *Startup firms active on social media are more likely to receive funding from more angel investors.*

Many investors concentrate their investments in a limited number of industries to better leverage specialized expertise, experience and business relationships that would help identify new opportunities (Sorenson and Stuart 2001). Investors making investments in a few specific industries should rely less on social media to discover new investment opportunities. On the other hand, for investors interested in making diverse investments across different business categories, it is unlikely that they would be able to sustain a significant base of contacts in each line of business to stay informed about potential startups to finance. Therefore, we should expect social media to play a larger role in the discovery of startups to invest in for investors with diverse portfolios.

Hypothesis 4: *Startup firms' activities on social media have less effect on investors making concentrated investments in certain industries.*

Hypothesis 5: *Startup firms active on social media are more likely to receive funding from more investors with diverse investment interests.*

Social Media as Information Channel to Evaluate Startup Quality

Beyond the search costs related to discovering investable startups, investors also have limited information on which to evaluate their investments in new ventures, a problem exacerbated by an asymmetric information between investors and entrepreneurs where entrepreneurs have an incentive to over-represent their quality. Investors engage in a complex information acquisition process (due diligence) to screen and evaluate investment opportunities; the ability to evaluate deals is a key performance differentiator among early stage investors (Gompers 1995, Van Nieuwerburgh and Veldkamp 2009).

Prior studies show that activities on social media can reveal information on firm quality to investors in online financial markets. Social media presence increases the success of crowdfunding activities (Agrawal et al. 2011). Similarly, more social media contacts ("friends") on a peer-to-peer lending platform increases the chance of reaching a funding target (Lin et al. 2013). In addition, early stage firms usually do not have a fully functional product or service ready for sale yet, so social media success could reveal the potential market size and customer reception of the product or service, thus foretelling the startup's chance of success. Anecdotal evidence also suggests that investors are increasingly evaluating metrics of social media presence, such as the number of followers on Twitter or other metrics of social media reach—as well as the sentiment of social media content (e.g. reviews, feedback) about a firm—when they make investment decisions.[25] An effective social media presence can serve as a signal of startup quality in an investor's evaluation process.

Geographic distance increases the difficulty for investors to obtain quality information on startups, since information circulates more freely between geographically proximate

[25] Wall Street Journal. "If You Look Good on Twitter, VCs May Take Notice". September 30, 2013. http://www.wsj.com/articles/SB10001424127887324659404578499702279196058

people and firms (Rosenthal and Strange 2004). This pattern also holds in the venture capital industry: VC investments are geographically concentrated (Sorenson and Stuart 2001), and 49% of all VC investments are made to startups located in the Boston, New York and San Francisco metropolitan areas (Chen et al. 2010). Startups located outside these VC clusters usually lack the face-to-face interaction channels to build reputation and trust as startups located closer to investors do, and it is also costly for investors to actually visit and inspect the startups located further away (Ivković and Weisbenner 2005, Lerner 1995, Massa and Simonov 2006). Therefore, investors have fewer channels of information to assess the quality of these startups outside VC clusters. In this case, we expect to see social media playing a more important role, providing information about these startups through the activities online and interactions with users, and having information easily accessible to investors. We hypothesize that:

Hypothesis 6: *Startups located further away from VC cluster regions see stronger effect of social media activities;*

The impact of social media activities also differs depending on investors' ability to process information. Experienced investors who have a substantial amount of previous investments, are likely to have accumulated expertise and knowledge that enable them to better process information and evaluate the quality of startup firms (Sørensen 2007). These investors should be more effective in analyzing the information from social media to evaluate startups and guide their investment decisions. Therefore, we expect that startups active on social media are more likely to obtain investment from experienced investors.

Hypothesis 7: *Startups active on social media are more likely to receive funding from experienced investors.*

On the other hand, if an investor already has other channels to learn about startups quality, this moderates the role of social media in providing quality information on startups. One major existing information channel for venture investors is their syndicate relationships. Venture capital firms and angel investors who jointly make VC funding in a given startup are called VC syndicates. Participating investors in the same syndicate usually have a substantial amount of interaction, as they coordinate investments and guide through startups' development. Thus, prior syndicates reveal close collaborative relationship between investors where information is shared (Hochberg et al. 2007). If investors have previous syndicate partners already invested in a startup, they can easily acquire information on startup quality from these partners, thus alternative quality information sources like social media is less important.

Hypothesis 8: *Startups' social media activities have less influence where investors have previous VC syndicate partners already invested in the firm.*

3.3 Data and Methodology

Sample Construction

The main dataset consists of investment rounds into new technology-based ventures in 2007–2015 obtained from CrunchBase, combined with data on startup social media activities on Twitter from Twitter API and Topsy.com.[26] Crunchbase records information on startups, people affiliated with the startups and investors. It focuses specifically on the information technology sector and has been considered to be representative of venture activity in their target markets (Block and Sandner 2009, Wu 2015). The Crunchbase data archive is obtained from a combination of user input and regulatory filings which are then reviewed for accuracy and compiled by TechCrunch staff. For each startup, there is data on the characteristics of each funding round to date (date, amount raised, type of funding, investor), characteristics of the venture itself (founding date, number of employees, type of business), and characteristics of the founders (prior venture experience and prior management experience).

We utilize Twitter as the source of social media data since it is the social media platform most extensively used by startups and investors, and broadly used by the business community; 60% of startups in our sample use Twitter while only 47% use Facebook and 36% used LinkedIn. The Twitter adoption rate for startups across different business categories are shown in Figure 3.1. We observe substantive Twitter usage by startups in different lines of business, with higher Twitter adoption rate in the news, media and information related industries, and lower adoption by transportation and manufacturing related businesses, as one would assume.

[26] Crunchbase is operated by TechCrunch, an AOL Inc. subsidiary delivering news on the information technology sector. Topsy.com is a certified Twitter partner, and maintains an archive of Twitter activity dating back to time Twitter was established (2006).

Some firms were excluded because the screenname utilized common English words (e.g., “path”, “square”, “tune”) which contaminate the data construction process on Twitter and Topsy.com, which rely on a text search of the firms’ screenname.

We focus specifically on the 2nd round of VC financing for three reasons. First, we do not want to use the 1st round of financing because not all startups are raising money, and we would not be able to empirically distinguish between those not raising money (“bootstrapping”) and those who are. Once a startup closes a 1st round of funding, it reveals that the firm is not bootstrapping, and consistent with the path of most technology startups backed by equity financing, they likely to need additional rounds of funding to sustain the firm. Second, we want to focus on earlier rounds of financing where public and private information available to investors is low and our theorized roles for social media still matters in reducing search costs and serving as a quality signal. In the later funding rounds, the theorized role of social media as an information channel would be harder to detect since there the startup firm has a track record already. Combining these first two points, the 2nd round is obviously the earliest round that isn’t the 1st round. Third, using the 2nd round allows us to use the 1st round as a control for firm size and quality.

Our data is primarily collected prior to the recent SEC regulation change in June 2015. In our observation window, startup firms have restrictions on the content they post on social media, specifically limiting the announcement of investment information to the public. We expect to see that after the regulation change, startup firms will more actively use social media to reach out to investors, and but exact empirical effect of social media on financing outcomes remains an open empirical question for future research.

Social Media Variables

We identify the Twitter page and screenname for each firm's corporate account (if it exists), and then use the Twitter API and Topsy.com API to gather information on Twitter activity, including:

-- **Number of tweets posted**: the number of distinct Tweets for each screenname that contain a link;[27]

-- **Mentions:** the number of distinct social media posts (tweets or links) that mention a startup's Twitter screenname in each month;

-- **Impressions:** the number of potential views of a firm's Tweets in each month;[28]

-- **Sentiment:** a normalized score from 0 (most negative) to 100 (most positive) based on the sentiment of all tweets mentioning a firm's screenname in each month;[29]

-- **Number of followers:** a count of the number of Twitter followers for each screenname.[30]

We include 3 measures of Twitter activity: 1) whether the firm created its Twitter account prior to receiving the 2nd round funding (***Started Using Twitter***), 2) the total number of Tweets posted in the 12-month-window prior to receiving the 2nd round funding (***Number of Tweets***), 3) the first principal component of the number of mentions, impressions, followers, and sentiment in tweets mentioning the Twitter account of the startup (***Twitter Influence*)**.

---

[27] Drawn from the Topsy.com archive, we utilize this proxy in lieu of the raw number of Tweets due to data limitations.
[28] The Impressions variable is provided by the Topsy.com API, calculated by multiplying the number of tweets mentioning the startup's by the number of followers during each month of our sample period.
[29] Sentiment was provided by the Topsy.com API.
[30] The number of followers was constructed for our dataset by taking a snapshot at a single point in time, namely June 12th, 2015 at 18:00.

Dependent Variables

We focus on funding outcomes as the dependent variable (closing a 2^{nd} round of funding, the number of investors participating in the 2^{nd} round, and the size of the 2^{nd} round funding) rather than other kinds of startup or VC performance measures (e.g. startups' successful exit; investors' returns to investment) because we are currently concerned with the link between social media and financial markets for early stage capital, although looking at other performance measures would be an excellent avenue for future research. The size of a funding round is a good general measure of fundraising outcomes, and since larger amounts raised are correlated with larger valuation, it also provides some insight on the investor's expectation of the startup's profitability and growth.

Control Variables

We include controls for startup characteristics (age, number of employees, and number of lines of business), founder characteristics (prior startup experience, prior executive-level management experience), industry (industry indicators), and year (year indicators) received the 2^{nd} round funding. The time and industry controls address market-wide conditions that could potentially affect funding. Overall, these variables control for variation in startup quality and are consistent with the prior literature on entrepreneurial financing (Hsu 2007).

To isolate the effects of Twitter from general online presence or other social media, we include controls for web site traffic rank (Alexa rank of a firm's homepage URL), search popularity (Google Trends data for a firm name as a search term), and an indicator for the firm's presence on Facebook. These variables also control for other marketing activity and brand awareness in addition to directly measuring online presence.

To control for communication from between prior investors in the startup to other investor through their personal contacts, we include a measure of investors' network connections through their syndicate partners. We use the PageRank measure to capture how well-connected the investors are and their ability to spread word about the startup to other investors; the PageRank measure captures the relative importance of nodes by factoring in how many connections they have and how important these connections are (Brin and Page 2012).

By including an extensive number of startup firm characteristics, including the size of the 1st round of financing, we control for many sources of unobserved firm quality that could potentially confound our estimates of social media's effect on funding success. Furthermore, since many of these variables are lagging indicators (prior year firm characteristics) or measures of changes (e.g. a firm adopting Twitter), we are less vulnerable to simultaneity between investment and social media use.

Regression Model

After log-transforming the round size measures, we estimate the following OLS model (with robust standard errors):

$$\log(2ndRoundFunding) = \beta_0 \log(1stRoundFunding) + \beta_1 Interval$$

$$+\boldsymbol{\beta_2 Started\ Using\ Twitter + \beta_3 NumberofTweets + \beta_4 Twitter\ Influence}$$

$$+\beta_5 WebsiteTrafficRank + \beta_6 GoogleTrend + \beta_7 OtherPlatform$$

$$+startup\_controls + founder\_controls + business\_category + year + \varepsilon$$

The model relates the amount raised in the 2nd round of financing to the amount raised in the 1st round of financing, the time elapsed between rounds and the Twitter activity measures.

We report the summary statistics and correlation between main variables of interest in Table 3.1 and Table 3.2. Our data includes 2,880 startup firms, for 2nd round funding events across years 2007-2015. The data is structured in a cross-sectional, with each startup firm appearing once. Social media measures are matched to the specific time-window before the 2nd round funding. Most of the other controls—such as firm age, website traffic, Google Trends measures, and founder controls—are matched to the specific timing of the round as well. However, our measures of number of employees and the number of followers on Twitter, with are fixed based upon our time of data collection, and the year indicators should address the natural time trend in these variables.

3.4 Results

Social Media Activities and General Funding Outcomes

To test our initial hypotheses that social media use is related to funding outcomes, we estimate Equation (1) for the full set of startups for which we have complete data using ordinary least squares (OLS). We first take a look at the overall influence of social media activities on startup funding outcomes, using data on startups' total amount of funding collected and the number of investors that they collect funding from in the 2nd VC funding round. In Table 3.3, we report the results relating the log value of total funding collected to the social media metrics and other control variables. The control variables all have signs in the right direction: startups who collected larger amount of funding in the 1st round, having more visits to their webpages (lower traffic rank) and attention from consumers (higher search volume as reported in Google Trend for query of startups company names) are also likely to collect more funding in the 2nd round; shorter interval between the two rounds are related to larger amount of 2nd round funding, as do startups with founders that worked on more startup projects previously and with more executive management experience, but the effect sizes are smaller in these cases.

Regarding the social media activity measures, we show that just being present on Twitter does not lead to startups' receiving larger amounts of funding. The number of tweets startups post on Twitter also has little effect on funding outcomes. In fact, posting too much information could actually have a negative effect, most likely due to the cost of managing tweets and lack of channel to really absorb the information collected from social media (Fischer and Reuber 2011). On the other hand, we are see strong positive effect of all the metrics relating to startups' influence on Twitter. Specifically, getting mentioned more in other people's tweets, have more impressions of tweets, with more positive sentiments in others' tweets mentioning the startup firm and a larger follower-base can all improve startups' funding outcomes. Since these measures are correlated with one another and show consistent results, we take their 1st principal component (Twitter Influence), to capture the overall impact (Column 1, Table 3.4). We find that a one standard deviation increase in the Twitter Influence measure leading to extra 1.1 million in 2nd round funding. This result is consistent if we use the subsample of startups who have started Twitter pages prior to receiving their 2^{nd} round funding (column 2). These results support our first hypothesis that social media activities improve startups' funding outcomes.

Next, we look at whether startups' activities on Twitter allow them to draw in a larger pool of potential investors. In columns (3) and (4) in Table 3.4, we use the total number of investors in the 2^{nd} VC funding round as the dependent variable, and found that startups with more influential social media profiles are likely to get more investors to make investments. This supports our second hypothesis that social media activities help startup firms get funded by a larger pool of investors. In column (5), we examine whether social media presence improves chances of getting 2^{nd} round funding in the first place, using the sample of all the startup firms that have received a 1^{st} round VC funding. We use a binary variable indicating whether the startup receives a 2^{nd} round funding as the dependent variable, and relate it to the social media measures, controlling for the size of the 1^{st} round funding and other startup and entrepreneur characteristics. Results show that startups' presence on Twitter and having high Twitter Influence can improve their chances of receiving 2^{nd} round funding.

The above results consistently demonstrate that startup firms' social media activities influence their funding outcomes. Our results suggest that startups should be effective in their social media activities to build a positive brand image, draw in a larger followers group, get more users to retweet their messages and have people leave more positive feedbacks relating to their business. Startups that are more successful at generating influence on social media see higher chances of continuing to receive funding, from a larger pool of investors and getting larger amounts of funding overall.

Social Media Activities and Discovery of Investment Opportunities

We have demonstrated so far that startups' social media activities contribute to funding success. In the next step, we turn to investigate the mechanisms of social media's influence on startup funding, through the discovery and evaluation of investment opportunities respectively. Firstly, we take a look at how social media presence influences investors' search for potential startup firms. We hypothesized that for investors with fewer channels of information to learn about potential investment opportunities, social media's function as a platform for broadcasting information is more important. We test for this by examining the composition of investors participating in startups' 2nd round funding, looking at the number of angel investors[31], while controlling for the total number of investors in the round[32]. In columns (1) and (2) of Table 3.5, we show that simply by having a Twitter presence, startups are more likely to have a larger portion of angel investors in the 2nd round funding, whereas the Number of Tweets and Twitter Influence have less influence. Since we are controlling for existing investors spreading word out about this startup by the PageRank measure of existing investors' VC syndicate connections, the result that more angel investors joining in the 2nd round for startups with Twitter accounts is most likely due to investors' discovery of new investment opportunities through social media.

Investors' own experience from previous investments and particularly investments in certain industries also build up connections that investors can refer to in order to learn about

---

[31] We are using the number of angel investors participating in the 2nd round funding here. Results are consistent if we use the number of all angel investors have not participated in previous rounds and only newly joined in the 2nd round; similar for the investors with diverse portfolios and investors with industry focuses.
[32] While angel investors generally invest in earlier stages of startups' development; it is not uncommon for angel investors to participate in the VC funding rounds as well.

new investment opportunities. Therefore, we expect to see social media as play a larger role in discovering startups for investors with more diverse investment portfolios. On the other hand, for investors making concentrated investments in certain industries and have consequently accumulated channels of information to learn about new investment opportunities, we expect to see social media playing a smaller role. To measure the diversity in investors' investment portfolios, we look at the investors' previous investments in other startups and the business categories they belong to. We define investors ranked in the upper 25th percentile of number of categories covered in previous investments as investors making diverse investments[33]. Investors with industry focuses as defined as those with total number of business categories covered in previous investments ranking in the lower 25th percentile.

Columns (3) and (4) in Table 3.5 show that startups active on social media are more likely to get more investors interested in making diversified investments to participate in the 2nd round funding. In contrast, columns (5) and (6) show that startups active on social media generally have a lower ratio of investors making investments in specific industries. Together, these two piece of evidence suggest that for investors investing in a business category they are familiar with, having sufficient connections with other investors and entrepreneurs to hear about new investment opportunities, social media's role of broadcasting information about startups and potential investment opportunities is less salient. On the other hand, for investors interested in making investments across multiple

[33] Similar results if we use the Herfindahl-Hirschman Index of previous investments across different categories to define diversity of investors' portfolio.

business categories, who are less likely to be master in all the categories, social media can be an effective channel of learning about startups in different lines of business and expanding the potential pool of investment opportunities.

These results are consistent with our hypotheses 3-5, showing that social media facilitates the entrepreneurial financing process, by providing information about startups, reducing the search cost and encouraging investors to explore a wider pool of startup firms, especially for investors with fewer channels of information, and for investors looking to make investments across different business categories but lack the industry connections to know about potential investment opportunities otherwise.

Social Media as Additional Information Channel for Startup

Once investors have identified the potential startup firms, the next step is for them to evaluate the investment opportunity and decide whether to actually fund each startup. We hypothesize that social media helps investors with this process, by providing more information about startup quality. For example, from startups' social media profiles, investors can learn about the startup's ability to build brand names through the online channel, reach out to target client groups, and also about consumers' feedback on the startups' products and services. Such additional information can help investors better evaluate the quality of the startup firms and make their investment decisions.

We first look at the funding outcomes for startups located outside VC clusters, i.e. outside the Boston, New York and San Francisco regions. These startups are located further away from investors and geographic distances can potentially exacerbate both the search cost and difficulty in obtaining quality information. While exact terms of startup financing and valuation are not publicly available, the size of the funding collected, which reflects investors' valuation of the startup firm, allows us to examine whether startups' social media activities influence investors' assessment of quality and valuation for the startups. In Table 3.6 we use the dummy variable (***Far from VC***) to indicate startup location outside VC clusters and include its interaction terms with the social media activity measures. Results show that while startups located outside the VC clusters in general receives less funding than startups located inside the VC cluster regions, they see additional gains in funding size from influential social media presence, with one standard deviation increase in Twitter Influence metrics adding 1 million more funding, compared with startups located inside VC clusters (column 3). This effect is not driven mainly by the discovery of investment opportunities (the coefficient for the interaction term between the regional dummy and started using Twitter is negative), suggesting that simply being present on Twitter is not sufficient to improve investors' valuation for the startup; active management of the social media presence and showing credible information through engagement with customers are necessary to improve investors' valuation for the startup. These findings suggest that for startups located further from VCs, where investors incur higher cost to obtain information, social media could present an additional information channel.

As a second piece of evidence for the role of social media in conveying information about startup quality, we look at startups' ability to reach out to the experienced investors. Hypothetically, if social media only works through the channel of discovering more investment opportunities, then startups should attract more average investors and more experienced investors in similar patterns, with their active social media presence. However, if the information on social media provide useful information on startup quality, then the experienced investors are more likely to effectively use the information in making their financing decisions. We examine this mechanism in Table 3.7, taking a look at how social media activities influence the number of experienced investors in the 2nd round funding, i.e. those who has made more than 100 investments up to date, while controlling for total number of investors in the round. Results are consistent with our hypothesis, showing that startups with more influence on Twitter get a higher portion of experienced investors (column 1), similarly if we look at the subsample of Twitter users only (column 2). These results suggest that startups more active on Twitter disproportionally attracts more experienced investors to invest in them, most likely because these investors are more capable in analyzing the information on social media to discover startup quality and make investment decisions accordingly.

On the other hand, if the investors already have trusted channels of information to learn about the quality of the startups, we expect to see the role of social media as an information channel to be of less significance. Specifically, we look at whether there are investors from previous funding rounds who are partners with investors in the 2nd round in the same VC syndicates for other projects. If so, investors in the 2nd round can obtain credible information about this startup from these syndicate partner investors and rely less on information from social media to deduce the quality about the startups. Evidence supports this hypothesis: in columns (3) and (4), we control for the percentage of investors in the 2nd round with partners from previous VC syndicates already invested in the same startup firm (***VC Syndicate***), and include its interactions with the social media measures. We observe that when a larger portion of the investors have alternative channels of learning about startup quality from previous syndicate partners, the effect of social media in presenting quality signal for startups and attracting experienced investors to join in is less significant. These results support our hypotheses 7 and 8, showing that social media not only act as a channel of broadcasting information about startups and letting investors discover the startups, but also provides investors with another information channel to learn about startups' quality and helping with their evaluation process.

3.5 Robustness Checks

One main endogeneity concern with the empirical analysis is that both social media activities and entrepreneurial financing could be influenced by the latent startup quality. We have controlled for some of this through the website traffic and Google Trend controls, measuring the general public's interest in the startup firms, accessing the startup homepages for product and service offerings or searching for the relevant information. In addition, we seek to reduce the effect of this type of endogeneity through the use of instrumental variables. Our identification strategy focuses specifically on the model which utilizes funding outcomes as the dependent variable, since that model is most likely to be affected by unobserved startup quality that might simultaneously influence social media influence. Using the same instruments in the other models yields similar outcomes to the OLS results for these as well.

We use the following three sets of instrumental variables. First, we use social media activities of other startups located in the same region. For each region and year combination, we look at the Twitter presence, number of tweets posted and Twitter influence measures respectively for other startups located in the same region. Twitter usage for firms located in the same region is likely to be influenced by similar factors, like the number of Twitter users in the region and users' propensity to interact with startups online but other firms' social media activities should not directly influence the startup's own funding outcomes. Second, we use social media activities of other startups that their investors previously invested in. If investors have different preferences of social media usage, this will lead to a correlation of social media use among firms; however, since investment amounts depend on firms specific factors they are unlikely to be correlated (especially since multiple investors tend to participate in the same investment round). Finally, we use a geographic measure of the awareness and use of Twitter using Google trends data on the search term "twitter" from 2007-2015 in each state. If startups in this region are more active on Twitter, we expect this to be reflected in the Google Trends, as consumers query for Twitter related information. This instrument should be correlated with the social media metrics of the startups, but not be directly linked to startup quality or funding outcomes.

Results from 2^{nd} stage of 2SLS regression, using these three sets of IVs to instrument for startups' starting the Twitter pages, number of tweets posted on Twitter and Twitter influence measures and using 2^{nd} round funding as the dependent variable, are reported in Table 3.8. We do not find evidence of weak instrument problems based on the usual tests for first stage predictive power ($F(65, 2694) = 38.11$, $p= 0.0001$). Since the instruments help tease out the effect due to better quality startups also more likely to be present on social media, we are able to better estimate the impact of social media on startup funding. Results are largely consistent with what we observed before: presence on Twitter improves startups' amount of funding collected. This effect is mainly driven by Twitter influence rather than Twitter activity. The economic size of the effects are comparable with those in Table 3.4 with one standard deviation increase in the Twitter Influence measure leading to about $1.27 M increase in next period funding. We also verified the other results using the instrument variables and got consistent results.

Another approach to control for the unobserved startup quality is to use a panel structure setup, with observations for startup-year combinations and calculating the total amount of funding the startup firms has collected up to date. In Table 3.9, we relate the log of total funding collected up to date, to the social media activities measures and the startup and entrepreneur level controls, including startup level fixed effects to capture unobserved quality (Columns 1 and 2). Results are largely consistent with before, indicating that startups present on social media, actively posting tweets and having high influence measure, are more likely to collected more funding across the years. Compared with results in Table 3.4, in the Fixed Effects regressions, being present on Twitter and tweeting information also positively contributes to funding outcomes. This is probably due to the accumulated effect over the years of heterogeneity across startup firms. In addition, we instrument for the Twitter activity measures on top of the Fixed Effects model, we continue to observe that startups with stronger influence on Twitter are more likely to collect larger sums of funding in total (Columns 3 and 4). The directions of the effects are consistent with before, while the scales are slightly higher compared with columns 1 and 2. This is likely due to the fact that we are already controlling for startup fixed effects and having many control variables in place, the marginal effects captured by IVs could be larger in scale. Still, the IV results indicate that we are not over-estimating the size of the effect.

3.6 Conclusion

We find that startup firms active on social media have higher chances of getting funded, receive larger amounts of funding and have a larger number of investors consistent with the idea that social media provides information that facilities venture funding. These effects are localized to social media influence rather than social media use. We further find these effects are larger for investors that might lack channels for discovering investments (angels, diversified investors), and that funding outcomes are improved in conditions where there is likely to be significant information asymmetry (ventures located outside VC clusters, investors lacking social network ties to get information about a startup). Thus, the gains associated with social media appear to be attributable to both an awareness effect, where investors can learn about a larger number of potential investments, and an uncertainty reduction effect where quality uncertainty is reduced in settings where alternative quality signals are less effective. These results are robust to various econometric methods (controls, instrumental variables) for accounting for the problems related to unobserved variation in startup quality.

Our results highlight the importance for early stage ventures to establish a presence on social media, especially for those investments where social media success can provide an indicator of their ability to attract and retain customers. However, even firms that are not in consumer facing industries can still benefit from expanding awareness among investors. Given that our data is primarily in a period when there were restrictions on social media activity that limited investment-related communications, recent legislative changes that now allow for greater information sharing on social media will likely increase the effect of social media on funding success. Our results also imply that while "cheap talk" in the form of Twitter posts does not have much influence on funding as would be expected, the ability to effectively engage readers in social media (influence) does matter suggesting benefits of even modest improvements in information availability in settings where there is considerable information asymmetry. While the use of extensive startup and social controls, contrasts within the data, and instrumental variables for addressing unobserved heterogeneity in startup quality does suggest the possibility that these effects are causal, in future work we hope to explore the specific communications more directly to gain a better understanding of how this information is communicated by looking at the specific content of social media interaction. Overall, we hope that this study and future related studies contribute to a better understanding of how the entrepreneurial financing market is changing due to the social media information and what startup firms should do to take advantage of the new opportunities.

3.7 References

Aggarwal R, Gopal R, Gupta A, Singh H (2012). Putting Money Where the Mouths Are: The Relation Between Venture Financing and Electronic Word-of-Mouth. Information Systems Research. 23(3-part-2) 976-992.

Agrawal A, Catalini C, Goldfarb A (2011). The Geography of Crowdfunding. SSRN Working Paper Series.

Aldrich HE, Fiol CM (1994). Fools rush in? The institutional context of industry creation. Academy of management review. 19(4) 645-670.

Amit R, Glosten L, Muller E (1990). Entrepreneurial ability, venture investments, and risk sharing. Management science. 36(10) 1233-1246.

Angst CM, Agarwal R, Sambamurthy V, Kelley K (2010). Social contagion and information technology diffusion: the adoption of electronic medical records in US hospitals. Management Science. 56(8) 1219-1241.

Aral S, Dellarocas C, Godes D (2013). Introduction to the special issue-social media and business transformation: A framework for research. Information Systems Research. 24(1) 3-13.

Aral S, Walker D (2011). Creating social contagion through viral product design: A randomized trial of peer influence in networks. Management Science. 57(9) 1623-1639.

Bakos JY, Brynjolfsson E (1993). Information technology, incentives, and the optimal number of suppliers. Journal of Management Information Systems 37-53.

Bapna R, Umyarov A (2015). Do Your Online Friends Make You Pay? A Randomized Field Experiment on Peer Influence in Online Social Networks. Management Science.

Bharadwaj A, El Sawy OA, Pavlou PA, Venkatraman N (2013). Digital business strategy: toward a next generation of insights. Mis Quarterly. 37(2) 471-482.

Block J, Sandner P (2009). What is the effect of the financial crisis on venture capital financing? Empirical evidence from US Internet start-ups. Venture Capital. 11(4) 295-309.

Brin S, Page L (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. Computer networks. 56(18) 3825-3833.

Chen C, Chellappa R (2009). Music sales and online user activity at social networking sites. Citeseer.

Chen H, De P, Hu YJ (2015). IT-Enabled Broadcasting in Social Media: An Empirical Study of Artists’ Activities and Music Sales. Information Systems Research. 26(3) 513-531.

Chen H, Gompers P, Kovner A, Lerner J (2010). Buy local? The geography of venture capital. Journal of Urban Economics. 67(1) 90-102.

Chevalier JA, Mayzlin D (2006). The effect of word of mouth on sales: Online book reviews. Journal of marketing research. 43(3) 345-354.

Chung S, Animesh A, Han K, Pinsonneault A (2014). Firms' Social Media Efforts, Consumer Behavior, and Firm Performance: Evidence from Facebook. Consumer Behavior, and Firm Performance: Evidence from Facebook (June 10, 2014).

Clemons EK, Row MC (1992). Information technology and industrial cooperation: the changing economics of coordination and ownership. Journal of Management Information Systems 9-28.

Dellarocas C, Zhang XM, Awad NF (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. Journal of Interactive marketing. 21(4) 23-45.

Dessein W (2005). Information and control in ventures and alliances. The Journal of Finance. 60(5) 2513-2549.

Eesley CE, Wu L (2015). Entrepreneurial Adaptation and Social Networks: Evidence from a Randomized Experiment on a MOOC Platform. Available at SSRN 2571777.

Fischer E, Reuber AR (2011). Social interaction via new social media:(How) can interactions on Twitter affect effectual thinking and behavior? Journal of business venturing. 26(1) 1-18.

Forman C, Ghose A, Wiesenfeld B (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. Information Systems Research. 19(3) 291-313.

Ghose A, Han SP (2011). An empirical analysis of user content generation and usage behavior on the mobile Internet. Management Science. 57(9) 1671-1691.

Goh K-Y, Heng C-S, Lin Z (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. Information Systems Research. 24(1) 88-107.

Gompers PA (1995). Optimal investment, monitoring, and the staging of venture capital. Journal of finance 1461-1489.

Greenwood BN, Gopal A (2015). Research Note—Tigerblood: Newspapers, Blogs, and the Founding of Information Technology Firms. Information Systems Research.

Gurbaxani V, Whang S (1991). The impact of information systems on organizations and markets. Communications of the ACM. 34(1) 59-73.

Hitt L, Jin F, Wu L (2015). Who Benefits More from Social Media: Evidence from Large-Sample Firm Value Analysis.

Hochberg YV, Ljungqvist A, Lu Y (2007). Whom you know matters: Venture capital networks and investment performance. The Journal of Finance. 62(1) 251-301.

Hong Y, Chen P-Y, Hitt LM (2014). Measuring Product Type with Dynamics of Online Review Variances: A Theoretical Model and the Empirical Applications. Available at SSRN 2422686.

Inderst R, Müller HM (2004). The effect of capital market characteristics on the value of start-up firms. Journal of Financial Economics. 72(2) 319-356.

Ivković Z, Weisbenner S (2005). Local does as local is: Information content of the geography of individual investors' common stock investments. The Journal of Finance. 60(1) 267-306.

Kaplan SN, Strömberg PE (2004). Characteristics, contracts, and actions: Evidence from venture capitalist analyses. The Journal of Finance. 59(5) 2177-2210.

Lee D, Hosanagar K, Nair H (2014). The effect of social media marketing content on consumer engagement: Evidence from facebook. Available at SSRN 2290802.

Lerner J (1995). Venture capitalists and the oversight of private firms. The Journal of Finance. 50(1) 301-318.

Li X, Hitt LM (2008). Self-selection and information role of online product reviews. Information Systems Research. 19(4) 456-474.

Li X, Wu L (2014). Herding and Social Media Word-of-Mouth: Evidence from Groupon. Available at SSRN 2264411.

Lin M, Prabhala NR, Viswanathan S (2013). Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. Management Science. 59(1) 17-35.

Lin M, Sias R, Wei Z (2015). “Smart Money”: Institutional Investors in Online Crowdfunding.

Luo X, Zhang J, Duan W (2013). Social media and firm equity value. Information Systems Research. 24(1) 146-163.

Malone TW, Yates J, Benjamin RI (1987). Electronic markets and electronic hierarchies. Communications of the ACM. 30(6) 484-497.

Massa M, Simonov A (2006). Hedging, familiarity and portfolio choice. Review of Financial Studies. 19(2) 633-685.

Rishika R, Kumar A, Janakiraman R, Bezawada R (2013). The effect of customers' social media participation on customer visit frequency and profitability: an empirical investigation. Information systems research. 24(1) 108-127.

Rosenthal SS, Strange WC (2004). Evidence on the nature and sources of agglomeration economies. Handbook of regional and urban economics. 4 2119-2171.

Saxenian A (1991). The origins and dynamics of production networks in Silicon Valley. Research policy. 20(5) 423-437.

Shane S, Cable D (2002). Network ties, reputation, and the financing of new ventures. Management Science. 48(3) 364-381.

Shriver SK, Nair HS, Hofstetter R (2013). Social ties and user-generated content: Evidence from an online social network. Management Science. 59(6) 1425-1443.

Sørensen M (2007). How smart is smart money? A two-sided matching model of Venture Capital. The Journal of Finance. 62(6) 2725-2762.

Sorenson O, Stuart TE (2001). Syndication networks and the spatial distribution of venture capital investments1. American journal of sociology. 106(6) 1546-1588.

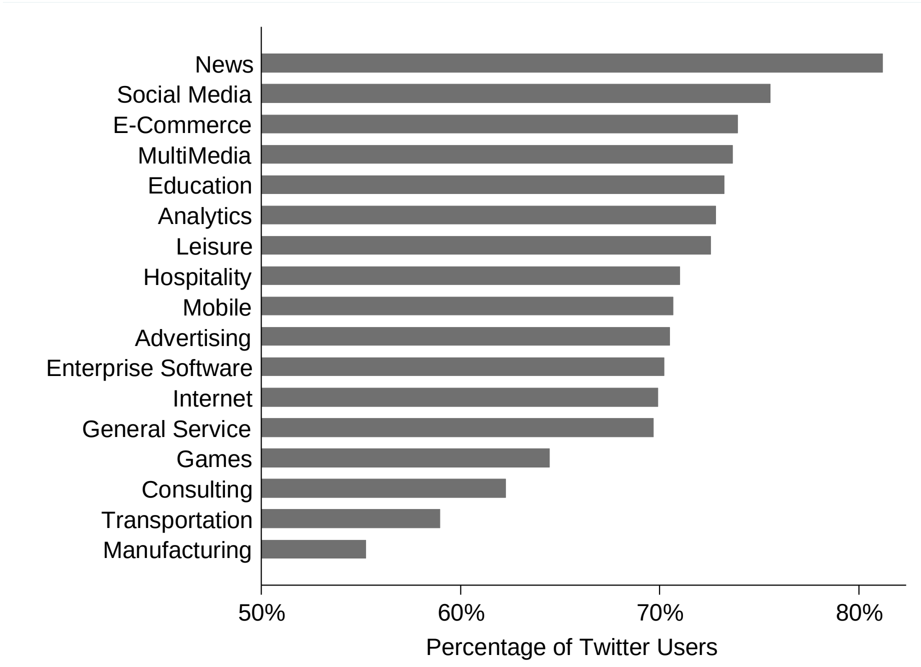
Van Nieuwerburgh S, Veldkamp L (2009). Information immobility and the home bias puzzle. The Journal of Finance. 64(3) 1187-1215.

Wu A (2015). Organizational Decision-Making and Information: Angel Investments by Venture Capital Partners. Available at SSRN 2656896.

Wu L, Jin F, Hitt LM (2014). Are All Spillovers Created Equal? A Network Perspective on IT Labor Movements. A Network Perspective on IT Labor Movements (November 20, 2014).

Zhu F, Zhang X (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. Journal of Marketing. 74(2) 133-148.

Figure 3.1. Twitter Adoption Rate across Startup Business Categories



Notes: 1. This graph shows the percentage of Twitter users for startups in different business categories; startups who has started a Twitter page by the time of our sample collection (June, 2015) are counted as Twitter users.

Table 3.1. Summary Statistics

Funding Round	**Obs.**	**Mean**	**Std. Dev.**	**Min**	**Max**
2^{nd} round funding	2,880	15,100,000	21,600,000	50,000	542,000,000
1^{st} round funding	2,880	6,994,867	8,027,593	48,268	124,000,000
Months between 1^{st} and 2^{nd} rounds	2,880	18.941	10.645	0	95
Year received 2^{nd} round funding	2,880	2011.072	2.628	2007	2015
Number of Investors in 2^{nd} Round	2,880	3.491	2.422	1	29
Firm Controls					
Website Traffic Rank	2,880	2,550,414	1,034,335	444	3,267,739
Google Trends	2,880	6.759	13.300	0	78.5
Startup Age	2,880	3.598	2.044	0	10
Number of Business Categories	2,880	2.582	1.994	1	14
Number of Employees	2,880	1451.358	10620.75	1	87673
Existing Investors' Page Rank	2,880	0	1	-1.181	3.305
Founder Controls					
Founders' Previous Projects	2,880	1.604	0.879	1	17
Founders' C-Level Experience	2,880	0.893	0.782	0	17
Twitter Measures					
Started Using Twitter	2,880	0.551	0.497	0	1
Number of Followers	2,880	13763.51	97573.78	0	2,439,962
Number of Tweets	2,880	265.153	747.298	0	12,914
Twitter Mentions	2,880	4262.136	29334.130	0	764,171
Sentiment	2,880	37.390	27.244	0	99
Impressions	2,880	160162.4	746651.4	0	9,397,203
Twitter Influence	2,880	0	1	-0.184	22.904

Table 3.2. Correlations between Main Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. 2^{nd} round funding	1							
2. 1^{st} round funding	0.3397	1						
3. Months between 1^{st} and 2^{nd} rounds	-0.0061	0.0233	1					
4. Website Traffic Rank	0.0101	0.0417	0.0591	1				
5. Google Trends	0.0382	0.0143	-0.0754	-0.1312	1			
6. Started Using Twitter	0.0057	-0.0278	-0.045	-0.0867	0.1342	1		
7. Number of Tweets	-0.0039	0.0091	-0.0003	-0.0781	0.1145	0.3201	1	
8. Twitter Influence	0.0085	-0.0013	-0.0287	-0.0384	0.1638	0.1658	0.2045	1

Table 3.3. Startup Social Media Activities and Funding

DV: **log(2nd round funding)** Label	(1) Base	(2) Started	(3) Tweets	(4) Mention	(5) Impressions	(6) Sentiment	(7) Followers	(8) All
log(1st round funding)	0.549***	0.549***	0.549***	0.547***	0.547***	0.550***	0.549***	0.547***
	(0.0261)	(0.0261)	(0.0261)	(0.0261)	(0.0259)	(0.0261)	(0.0261)	(0.0259)
Interval between Rounds	-0.00216	-0.00216	-0.00214	-0.00218	-0.00193	-0.00218	-0.00214	-0.00188
	(0.00163)	(0.00163)	(0.00163)	(0.00163)	(0.00161)	(0.00163)	(0.00164)	(0.00161)
Website Traffic Rank	-0.00725	-0.00730	-0.00715	-0.00207	-0.00283	-0.00245	-0.00687	0.000158
	(0.0264)	(0.0264)	(0.0264)	(0.0264)	(0.0261)	(0.0264)	(0.0264)	(0.0262)
Google Trends	0.00333***	0.00335***	0.00341***	0.00290***	0.00335***	0.00305***	0.00324***	0.00321***
	(0.00104)	(0.00104)	(0.00104)	(0.00104)	(0.00104)	(0.00105)	(0.00104)	(0.00105)
Has Facebook Page	0.0865***	0.0901***	0.0911***	0.0799**	0.0723**	0.0847**	0.0897***	0.0698**
	(0.0334)	(0.0340)	(0.0341)	(0.0342)	(0.0341)	(0.0341)	(0.0341)	(0.0342)
Existing Investors' Page Rank	0.267***	0.266***	0.268***	0.286***	0.270***	0.265**	0.267***	0.289***
	(0.0978)	(0.0989)	(0.0984)	(0.0979)	(0.102)	(0.106)	(0.0999)	(0.0961)
Started Using Twitter		-0.0174	-0.0154	-0.0102	-0.131***	-0.0516	-0.0146	-0.107**
		(0.0426)	(0.0427)	(0.0426)	(0.0458)	(0.0444)	(0.0425)	(0.0466)
Number of Tweets			-0.00264					-0.00687*
			(0.00406)					(0.00392)
Twitter Mention				0.0870***				0.0573***
				(0.0209)				(0.0222)
Impressions					0.132***			0.124***
					(0.0228)			(0.0261)
Sentiment						0.0492***		-0.00833
						(0.0173)		(0.0192)
Number of Followers							0.180*	0.0985
							(0.0922)	(0.0963)
Other Controls	Year Receiving Funding, Business Category, Firm Age, Number of Employees Number of Previous Venture Projects by Founders, Founders' C-level experience, other funding							
Constant	7.285***	7.293***	7.286***	7.313***	7.388***	7.314***	7.297***	7.377***
	(0.401)	(0.402)	(0.402)	(0.401)	(0.401)	(0.404)	(0.402)	(0.399)
Observations	2,880	2,880	2,880	2,880	2,880	2,880	2,880	2,880
R-squared	0.292	0.292	0.292	0.296	0.300	0.294	0.292	0.302

Notes: 1. Dependent Variable is the log value of total amount of funding collected;
2. Reporting results using OLS regression; *** p<0.01, ** p<0.05, * p<0.1

Table 3.4. Startup Social Media Activities and Funding Outcomes

	(1)	(2)	(3)	(4)	(5)
Dependent Variables	**Log(2^{nd} Round Funding)**		**Number of Investors**		**Getting 2^{nd} Round Funding**
Sample	All	Twitter	All	Twitter	All
log(1st round funding)	0.547***	0.536***	0.328***	0.245***	0.144***
	(0.0260)	(0.0282)	(0.0648)	(0.0939)	(0.0557)
Interval between Rounds	-0.00194	-0.00522**	-0.00628	-0.00695	
	(0.00162)	(0.00209)	(0.00420)	(0.00591)	
Existing Investors' PageRank	0.290***	0.359***	-0.389	-0.143	0.446
	(0.101)	(0.0615)	(0.326)	(0.381)	(1.913)
Started Using Twitter	-0.0812*		-0.316***		0.570***
	(0.0442)		(0.117)		(0.134)
Number of Tweets	-0.00567***	-0.00598***	-0.000235	-0.00233	-0.000686
	(0.00197)	(0.00204)	(0.00731)	(0.00762)	(0.00559)
Twitter Influence	0.115***	0.112***	0.240***	0.188**	0.192**
	(0.0193)	(0.0215)	(0.0671)	(0.0757)	(0.0795)
Other Controls	Year Receiving Funding, Business Category, Firm Age, Number of Employees Number of Previous Venture Projects by Founders, Founders' C-level experience, Website Traffic Rank, Google Trends, Facebook Presence, other funding				
Constant	7.387***	7.231***	-2.345**	-2.903*	4.267
	(0.402)	(0.515)	(1.083)	(1.605)	(3.834)
Observations	2,880	1,588	2,880	1,588	6,378
R-squared	0.300	0.369	0.159	0.196	0.198

Notes: 1. In columns (1) and (2), the dependent variable is the log of 2nd round funding; in columns (3) and (4), the dependent variable is the number of investors in the 2nd round; in columns (5) and (6) the dependent variable is a 0/1 indicating whether the startup firm has received 2nd round funding, using all firms that received 1st round funding as the sample of analysis, reporting results from logistic regression

2. Columns (1)(3)(5) uses all startups and columns (2)(4) use the subsample of startup that has started Twitter page

3. *** p<0.01, ** p<0.05, * p<0.1

Table 3.5. Social Media and Discovering New Investment Opportunities

Dependent Variables	(1) Number of Angel Investors	(2)	(3) Number of Investors with Diverse Portfolios	(4)	(5) Number of Investors with Industry Focus	(6)
log(1st round funding)	-0.118***	-0.117***	0.238***	0.237***	0.0424	0.0431*
	(0.0239)	(0.0238)	(0.0287)	(0.0285)	(0.0261)	(0.0261)
Interval between Rounds	0.000271	0.000270	-0.00789***	-0.00775***	0.000123	7.22e-05
	(0.00140)	(0.00140)	(0.00194)	(0.00194)	(0.00172)	(0.00171)
Existing Investors' PageRank	0.0701**	0.0719**	-0.0738	-0.0424	0.139	0.126
	(0.0353)	(0.0338)	(0.233)	(0.224)	(0.277)	(0.273)
Number of Investors	0.216***	0.215***	0.249***	0.245***	0.178***	0.180***
	(0.0237)	(0.0238)	(0.0159)	(0.0157)	(0.0177)	(0.0177)
Started Using Twitter	0.176***	0.139***	0.170***	0.0474	-0.187***	-0.136***
	(0.0320)	(0.0364)	(0.0558)	(0.0582)	(0.0471)	(0.0497)
Number of Tweets		0.00248		0.000121		-0.000417
		(0.00318)		(0.00338)		(0.00154)
Twitter Influence		0.0325*		0.174***		-0.0709***
		(0.0195)		(0.0281)		(0.0195)
Other Controls	Year Receiving Funding, Business Category, Firm Age, Number of Employees Number of Previous Venture Projects by Founders, Founders' C-level experience, Website Traffic Rank, Google Trends, Facebook Presence, other funding					
Constant	1.118***	1.127***	-3.079***	-2.938***	-1.319***	-1.376***
	(0.334)	(0.333)	(0.483)	(0.479)	(0.446)	(0.445)
Observations	2,880	2,880	2,880	2,880	2,880	2,880
R-squared	0.319	0.320	0.429	0.438	0.440	0.443

Notes: 1. In columns (1) and (2), the dependent variable is the number of angel investors in the 2nd funding round; in columns (3) and (4), the dependent variable is the number of investors with diverse investment portfolios (i.e. in the upper 25th percentile of total number of business categories covered in previous investments); in columns (5) and (6) the dependent variable is the number of investors with industry focuses (i.e. invested in the business category before and in the lower 25th percentile of total number of business categories covered by previous investments)
2. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3.6. Startup Location, Social Media and Evaluation of Startups

DV: Log(2^{nd} Round Funding)	(1)	(2)	(3)	(4)
Label	Base	Region	Cross Effect	Cross Effect
Sample	All	All	All	Twitter Users
log(1^{st} round funding)	0.548***	0.537***	0.536***	0.521***
	(0.0260)	(0.0259)	(0.0260)	(0.0283)
Interval between Rounds	-0.00200	-0.00188	-0.00182	-0.00505**
	(0.00162)	(0.00161)	(0.00160)	(0.00207)
Existing Investors' PageRank	0.291***	0.305***	0.302***	0.364***
	(0.102)	(0.106)	(0.108)	(0.0788)
Started Using Twitter	-0.0856*	-0.0867**	-0.00429	
	(0.0442)	(0.0438)	(0.0516)	
Number of Tweets	-0.00660*	-0.00666*	-0.0109***	-0.0110***
	(0.00393)	(0.00390)	(0.00383)	(0.00381)
Twitter Influence	0.105***	0.0972***	0.0624***	0.0544**
	(0.0186)	(0.0186)	(0.0197)	(0.0217)
Far from VC		-0.172***	-0.0898*	-0.278***
		(0.0295)	(0.0516)	(0.0442)
Far from VC * Started Using Twitter			-0.166**	
			(0.0720)	
Far from VC * Number of Tweets			0.0146	0.0131
			(0.0130)	(0.0128)
Far from VC * Twitter Influence			0.0864**	0.106***
			(0.0376)	(0.0396)
Other Controls	Year Receiving Funding, Business Category, Firm Age, Number of Employees, Number of Previous Venture Projects by Founders, Founders' C-level experience, Website Traffic Rank, Google Trends, Facebook Presence, other funding			
Constant	7.365***	7.524***	7.487***	7.529***
	(0.402)	(0.402)	(0.403)	(0.513)
Observations	2,880	2,880	2,880	1,588
R-squared	0.299	0.307	0.310	0.386

Notes: 1. Far from VC is a binary variable indicating whether the startup firm is located within the VC cluster regions of Boston, New York and San Francisco.
2. Columns (1)-(3) uses all the sample and column (4) uses the subsample of startup that has started Twitter page before receiving 2nd round funding
3. *** p<0.01, ** p<0.05, * p<0.1

Table 3.7. Social Media, Quality Signals and Prestigious Investors

DV: Number of Prestigious Investors	(1)	(2)	(3)	(4)
Sample	All	Twitter Users	All	Twitter Users
log(1st round funding)	0.184***	0.259***	0.181***	0.254***
	(0.0211)	(0.0326)	(0.0212)	(0.0327)
Interval between Rounds	-0.00464***	-0.00527***	-0.00444***	-0.00486**
	(0.00137)	(0.00191)	(0.00139)	(0.00195)
Number of Investors	0.110***	0.125***	0.108***	0.123***
	(0.0102)	(0.0153)	(0.0104)	(0.0154)
Started Using Twitter	-0.00151		0.00283	
	(0.0395)		(0.0396)	
Number of Tweets	-0.00150	-0.00271	-0.00140	-0.00271
	(0.00243)	(0.00257)	(0.00239)	(0.00252)
Twitter Influence	0.147***	0.135***	0.142***	0.133***
	(0.0213)	(0.0248)	(0.0209)	(0.0244)
VC Syndicate			-0.0141	-0.0526***
			(0.0114)	(0.0142)
VC Syndicate * Started Using Twitter			-0.0549***	
			(0.0163)	
VC Syndicate * Number of Tweets			0.00508	-0.00345
			(0.0107)	(0.0126)
VC Syndicate * Twitter Influence			-0.0238**	-0.0166
			(0.0113)	(0.0133)
Other Controls	Year Receiving Funding, Business Category, other funding, Firm Age, Number of Employees, Number of Previous Venture Projects by Founders, Founders' C-level experience, Website Traffic Rank, Google Trends, Facebook Presence, Existing Investors' PageRank			
Constant	-2.833***	-4.111***	-2.829***	-4.118***
	(0.334)	(0.514)	(0.335)	(0.514)
Observations	2,880	1,588	2,880	1,588
R-squared	0.347	0.351	0.350	0.354

Notes: 1. The dependent variable is the number of experienced investors in the 2nd round, defined as those having made more than 100 investments up to date;
2. VC Syndicate is the percentage of investors in the 2nd round with partners from previous VC syndicates already invested in the same startup firm;
3. Columns (1) and (3) use all the sample and columns (2) and (4) use the sample of startups with Twitter page prior to receiving 2nd round funding;
4. *** p<0.01, ** p<0.05, * p<0.1

Table 3.8. Social Media Activities and 2^{nd} Round Funding, Using IVs

DV: log(2^{nd} round funding)	(1)	(2)
Sample	All	Twitter Users
log(1st round funding)	0.551***	0.524***
	(0.0238)	(0.0264)
Interval between Rounds	-0.00140	-0.00463**
	(0.00157)	(0.00181)
Existing Investors' PageRank	0.370	0.409*
	(0.236)	(0.217)
Started Using Twitter	0.508	
	(0.533)	
Number of Tweets	-0.0146***	-0.0114***
	(0.00481)	(0.00340)
Twitter Influence	0.235***	0.244***
	(0.0487)	(0.0426)
Other Controls	Year Receiving Funding, Business Category, Firm Age, Number of Employees, Number of Previous Venture Projects by Founders, Founders' C-level experience, Website Traffic Rank, Google Trends, Facebook Presence, other funding	
Constant	7.027***	7.299***
	(0.697)	(0.461)
Observations	2,762	1,527
R-squared	0.228	0.347

Notes: 1. Three sets of instruments used as instruments: 1) Google Trend for the keyword "Twitter" in each region; 2) Twitter usage, number of tweets and twitter influence in the other startups located in the same region; 3) twitter usage, number of tweets and twitter influence in other firms that the investors previously invested in.
2. Reporting the 2nd stage results from 2SLS regression
3. Columns (1) uses all the sample and column (2) uses the subsample of startups that started Twitter page before receiving 2nd round funding
4. *** p<0.01, ** p<0.05, * p<0.1

Table 3.9. Social Media Activities and Total Funding, Fixed Effects

DV: log(total funding)	(1)	(2)	(3)	(4)
Model	FE	FE	FE/IV	FE/IV
Sample	All	Twitter Users	All	Twitter Users
Firm Age	0.321***	0.404***	0.508***	0.368***
	(0.0143)	(0.0175)	(0.0292)	(0.0926)
Website Traffic Rank	-0.507***	-0.469***	-0.286***	0.716
	(0.0369)	(0.0397)	(0.0731)	(0.780)
Google Trends	-0.00570	-0.00885**	-0.0164	-0.0468**
	(0.00375)	(0.00386)	(0.0106)	(0.0231)
Started Using Twitter	1.340***	1.683***	-5.869***	17.26
	(0.0605)	(0.0673)	(1.046)	(21.86)
Number of Tweets	0.0219***	0.0218***	0.978***	2.348***
	(0.00591)	(0.00589)	(0.297)	(0.860)
Twitter Influence	0.766***	0.801***	2.100***	3.178*
	(0.0307)	(0.0309)	(0.520)	(1.768)
Other Controls	Year Controls, Business Category, Dummies for missing Variables			
Constant	11.36***	10.39***	7.497***	-9.711
	(0.542)	(0.581)	(1.019)	(13.25)
Observations	105,292	74,283	104,834	74,011
R-squared	0.738	0.739	18,054	13,257

Notes: 1. Dependent variable is the log of the total amount of funding collected up to date;

2. Columns (1) and (2) show results using fixed effects regression; columns (3) and (4) use fixed effects regression with instrumental variables. The instruments are: 1) Google Trend for the keyword "Twitter" in each region; 2) Twitter activities in the other startups located in the same region; 3) Twitter activities in other firms that the investors previously invested in.

3. Columns (1) and (3) uses all the sample and columns (2) and (4) use the subsample of startups that have eventually started Twitter Page

4. Robust standard errors reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1