The Emerging Geography of the Blockchain Industry

by

Martin Holicka

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Environmental Studies
in
Geography

Waterloo, Ontario, Canada, 2021

Author’s Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Geographers have long been interested in where new technologies and industries emerge. The presence and adoption of a new technology within a region has multiple positive and negative externalities. Scholars have commented on the contribution of these firms to the regional economy, in terms of increasing human capital, innovation, and research and development. Technology firms in particular tend to locate in world cities and technology hubs, with concentrations of highly skilled workers, venture capital, anchor institutions and knowledge infrastructure. Using the blockchain industry as a case, this thesis examines the geography of nascent industries. Blockchain, which emerged in 2009 and is best known for applications such as bitcoin, has application in supply chain optimization, royalty and copyright tracking, cybersecurity, refugee identity and transaction systems, and voting systems. Blockchain's widespread application across industries and regions provides an excellent opportunity to explore the emerging geography of tech firms. This study explores this geography and attempts to identify key patterns and locations. Using economic data from Crunchbase and analysis using Elasticsearch, this study demonstrates that blockchain firms follows similar patterns seen elsewhere in the tech industry. Large world cities remain at the forefront of both firm and investor activity, and they are shown to be of crucial importance in global networks. Based on these findings, the study concludes by encouraging policy makers to understand the importance of these key geographies and identifies areas for further research to advance our understanding.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Tara Vinodrai for all her guidance and support throughout this work. Through discussions with her, I was able to step back from the data and think critically about what conclusions may be derived, and this has been instrumental in writing the thesis.

I would also like to thank my defense committee and thesis readers, Dr. Markus Moos and Dr. Heather Hall for taking the time to read my thesis and provide me with their insights and feedback.

I also want to thank my family and in particular, my sister Caroline who helped me with a large number of edits, making the thesis flow and sound much better than the drafts.

Lastly, I would like to acknowledge funding and support from the Social Science and Humanities Research Council of Canada, as well as the Innovation Policy Lab at the Munk School of Global Affairs and Public Policy, University of Toronto.

Table of Contents

Author's Declaration ii
Abstract iii
Acknowledgements iv
List of Figures vii
List of Tables viii
Chapter 1 Introduction 1
Chapter 2 Literature Review 8
2.1 Entrepreneurial Ecosystems and Innovation 9
2.2 Clusters and The Geography of Innovation 13
2.3 Venture Capital and the Geography of Investment 20
2.4 Summary 23
Chapter 3 Data and Methods 24
3.1 Dataset Sourcing 25
3.2 Methods Overview 26
3.3 Database Acquisition and Cleaning 26
3.4 Elasticsearch 32
3.5 Data Analysis 34
3.5.1 Exploratory Analysis and Categorization 34
3.5.2 Network Analysis 38
3.6 Limitations 40
Chapter 4 Results and Discussion 43
4.1 Nature of the Blockchain Industry 43
4.1.1 Firms in the Industry 44
4.1.2 Firms Receiving Investment Funds 50
4.1.3 Investment by Firm Specialization 52
4.2 Geographic Patterns of the Industry 56
4.2.1 Global Geography of the Blockchain Industry 56
4.2.2 Geography of Investors 67
4.2.3 Geography of Firms 68
4.3 Geographic Patterns of Investment 70
4.3.1 Geographic Spread of Invested Firms 71

4.3.2 Network Analysis 73
4.4 Summary 79
Chapter 5 Conclusion 82
Bibliography 90
Appendix 111
Appendix A Aggregation into Silicon Valley 112
Appendix B Categorization Mappings 113
Appendix C Featured Graphic Submitted to Environment and Planning A 115

List of Figures

Figure 2-1 Global superstar cities..19
Figure 3-1 Database Acquisition and Cleaning methodology overview 27
Figure 3-2 Visual categorization methodology 37
Figure 4-1 Firm founding per year 45
Figure 4-2 Employee counts..46
Figure 4-3 Country-level activity (Firms and Investors)...57
Figure 4-4 Country-level firm activity 58
Figure 4-5 Country-level investor activity 59
Figure 4-6 Regional activity - North America..65
Figure 4-7 Regional Activity - Europe...65
Figure 4-8 Regional Activity - Asia and Oceania 66
Figure 4-9 The global geography of investment into blockchain firms 74
Figure 4-10 Eigenvector centrality...77

List of Tables

Table 3-1 Data Overview – Primary role of organization 28
Table 3-2 Data overview - investor type 29
Table 3-3 Elasticsearch vs RDBMS Terms 32
Table 3-4 Tier list of categories 36
Table 4-1 Application Area 48
Table 4-2 Firm status 49
Table 4-3 General vs Investment Representation 50
Table 4-4 Firms by Employee Count 51
Table 4-5 Investor and Firm category 53
Table 4-6 Country level firm counts 60
Table 4-7 Country level investor counts 60
Table 4-8 Investor activity by region 67
Table 4-9 Top city region firm activity 69
Table 4-10 Levels of firm investment (%) by country 71
Table 4-11 Betweenness Centrality 75

Chapter 1
Introduction

Geographers have long been interested in where new technologies and industries emerge. The presence and adoption of a new technology within a region can have both positive and negative externalities. Scholars have commented on the contribution of these firms to the regional economy, in terms of increasing human capital, innovation, and research and development, in addition to the presence of entrepreneurial individuals (Bollinger et al., 1983). Scholars have noted that some regions are far more successful than others in attracting firms and individuals (Florida, 2014; Saxenian, 1994). The reason posited is that some locations are more suitable for certain economic activity due to the presence of actors and assets within the locale that promote the production or consumption of goods and services, all within close geographic proximity (Mudambi et al., 2018). These ideas have been examined through using several theoretical concepts, such as industrial districts (Amin, 1989) and spatial clusters (Bathelt et al., 2004a). The main advantages of co-location are the presence of financial and human capital, access to global knowledge pipelines (Bathelt et al., 2004a), and a drive to innovate and remain competitive, as there is a high penalty for falling behind co-located peers (Porter, 2006). The regions that best fit these descriptions are typically world cities. These locales are able to provide both higher wages and the needed infrastructure and ecosystem for firms (Derudder, 2008). However, the concentration of the technology industry within a region can also been associated with growing social and spatial inequality, leading to issues related to exclusion, housing affordability, gentrification and polarization within the labour market (Beaulieu et al., 2004; Brail & Vinodrai, 2020; Florida, 2017; Glaeser et al., 2009)

Firms within the tech industry in particular tend to locate in world cities and technology hubs, with concentrations of highly skilled workers, venture capital, anchor institutions and knowledge infrastructure (Brail, 2020)[1]. Higher wages and the quality of the overall innovation ecosystem can be factors that attract highly skilled workers. The main advantages for a worker in

[1] A further discussion regarding the usage of "world cities" and the different classifications may be found in Section 2.2.

the region include access to specialized training programs, access to tacit knowledge and the ability to network with like-minded individuals. Additionally, high quality, supportive institutions such as universities and specialized economic development policies that aim to bolster innovation in the region can add to the appeal of particular locations (Florida, 2017; Florida & Hathaway, 2018; Lee & Clarke, 2019). In addition to a highly specialized worker pool, access to venture capital and funding is a large draw for these firms. Scholarship on entrepreneurial ecosystems and start-ups indicates that venture capital firms will locate in areas with access to financial capital, combined with access to important global networks (Mingo et al., 2018). In this way, these firms are able to source the capital needed to fund their portfolio and expand their portfolio by tapping into the network of other investors, either through investor conglomerates or networking. For example, an investor may learn about an up-and-coming firm through their network and be able incorporate that firm into their portfolio. Additionally, geographic proximity of the firm assets within the portfolio has been shown to increase portfolio firm performance, both in terms of exits and overall funding (Chen et al., 2010; Kang et al., 2020).

Firms within the tech industry tend to co-locate within world cities partially due to the presence of anchor institutions. These institutions support the entrepreneurial ecosystem and provide some of the necessary infrastructure for firms to thrive. Such institutions are recognized as leaders and facilitate both the attraction of highly skilled workers and pipelines to outside knowledge sources. In the case of established world cities, anchor institutions can be thought of as government and educational institutions that enhance place-based competitive advantages. This may be accomplished by specialized funding through regional policy, or the presence of a world-renowned educational or research institution, which facilitate innovation and knowledge transfer to the region (Breznitz, 2014; Gertler & Vinodrai, 2005). A subset of this scholarship focuses on the role of anchor institutions in secondary and tertiary cities. These cities are supported by innovation policy with the aim of attracting an anchor firm, typically through governmental subsidies for innovation, military spending (Senor & Singer, 2011) or attractive tax incentives (Brail, 2019).

The establishment of an anchor firm motivates the development of surrounding support systems, which drives innovation and investment within the region. For example, consider a tech giant such as Apple locating in a region. The presence of the firm will not only attract highly skilled workers from other regions but can also facilitate the growth of a cluster due to the co-location of firms from surrounding regions. Firm relocation to these secondary and tertiary cities may also be a result of the optimization of costs for firms and their workers. From a firm perspective, a decrease in taxes, property and wages may be a factor in the decision to relocate (Glaeser et al., 2009). Costs for workers such as housing, and overall quality of life for the employees may also be a cause for relocation (Brail & Vinodrai, 2020; Florida, 2019; Lee, 2019). Co-location with anchor firms in these secondary and tertiary cities can allow firms to have access to a subset of the same externalities gained from locating in a world city, potentially at a lower cost (Glaeser et al., 2009).

An analysis of emerging technology firms offers an opportunity to contribute to current debates about the locational dynamics of the tech industry. Blockchain is one such emerging technology. Blockchain is a distributed ledger technology that maintains transparency, security, and accountability in transactions. It was first described in 2008 in a whitepaper entitled *Bitcoin: A peer-to-peer electronic cash system,* which was published by the unknown individual / group known only as Satoshi Nakamoto. This publication was followed by a software release in January 2009, which created a technology aimed to address the need for third-party trust. It did so by distributing the knowledge of each transaction across the system and forcing the agreement of each transaction within the system. The system works by requiring the confirmation by a peer of a transaction that occurred through solving complex mathematical and cryptographical equations. In this way, no one peer controls the system, and thus this forces transparency (Nakamoto, 2009). Indeed, transparency is a key characteristic of blockchain technology, alongside robustness, auditability and security (Christidis & Devetsikiotis, 2016). These characteristics have made blockchain technology of high interest to firms and other organizations as a means to make their transactions safer and more transparent.

While Nakamoto introduced blockchain for use in decentralized, anonymous currencies, the technology has had wide application in a variety of industries. Blockchain has already seen numerous applications across industries, such as food supply chains (Tan et al., 2018), financial loans (Yang et al., 2018), elections (Barnes et al., n.d.; Kshetri & Voas, 2018), education (Chen et al., 2018), health (Angraal et al., 2017; Drosatos & Kaldoudi, 2019), and energy (Cheng et al., 2017). An example of a detailed application is the usage of transaction tracking to establish the drug verification programs, in an effort to limit drug fraud, which has been found to have a large impact on the industry (US Food and Drug Administration, 2019)[2]. Each application of the technology has the potential to revolutionize or 'disrupt' current practices. Due to the broad use of blockchain from financial and service-based applications such as banks, marketing and cryptocurrencies to industrial uses such as smart contracts, blockchain technologies have garnered the interest of business and industry leaders around the world (IBM, 2017). Indeed, a wide range of leading consulting companies, global think tanks, industry observers and scholars promote blockchain as a potential industry-changing or 'disruptive' technology (Carson et al., 2018; Deloitte, 2020; PricewaterhouseCoopers, 2018). Recent reports suggest that the technology is no longer about early adopters and experiments, but rather, "robust enterprise ready solutions" (Deloitte, 2020, p. 3). These reports point to the fact that blockchain is an enabling technology with wide application.

Recent scholarship on the blockchain industry has examined the role of cryptocurrencies in financialization, the relationship between venture capital and start-up dynamics, and firm competitiveness (Fernandez-Vazquez et al., 2019; Park et al., 2020; Zook and Grote, 2020). However, few studies focus explicitly on the geography of the blockchain industry or the investment flows supporting these firms. Most scholarly studies of the industry are either technically focused explorations or economic case studies about the usage and power of blockchain utilization in corporations. More technical papers examine the technology through the lens of security and technical usability (Dinh et al., 2018; Pierro, 2017; Zhang et al., 2019) along with cryptocurrency applications (Bartoletti et al., 2019; Wu et al., 2019). Elsewhere, scholars have

[2] More detailed explanations of blockchain usage is described in Section 4.1.3

examined the business use of blockchain (Gatteschi et al., 2018; Tapscott & Tapscott, 2017), industry and investment dynamics with a focus on firms that changed their names to add blockchain or bitcoin as a way to capitalize on the surge of investments (Jain & Jain, 2019), and specific applications for blockchain through the lens of smart cities (Sun et al., 2016). While these studies add to the conversation about blockchain in a unique way, none of these papers explore the blockchain industry as a whole.

Recent papers regarding the investment geography of the blockchain industry have discussed the industry from the perspective of industrial application within the finance sector. Additionally, they examine the flow of venture capital into various application areas, concluding that venture capital investments remain focused on finance, insurance and information and communication applications (Fernandez-Vazquez et al., 2019; Zook & Blankenship, 2018). A study by Friedlmaier et al. (2018) examines the regional distribution of blockchain firms and identifies the presence of two leading geographical regions: the US and UK. None of these papers offer a broader global analysis of the industry and our understanding of these global patterns and dynamics remains limited. It remains an open question as to whether the blockchain industry follows well-established patterns associated with other technology industries. This relates to a central debate amongst economic geographers and other scholars regarding the geographic conditions and patterns of emerging or new industries and clusters (Boschma, 1997; Braunerhjelm & Feldman, 2006; Heiberg et al., 2020; Henn & Bathelt, 2018). However, it is expected that the blockchain industry will not deviate – in aggregate - from these patterns, which have been documented in other technology industries, as it can be argued that these emerging tech firms require similar ecosystems and will profit from the same local externalities described above.

To help address this gap in our knowledge, this thesis explores the global geography of the blockchain industry. More specifically, the aim of this research is to understand the firm and industry dynamics associated with blockchain activity within a *global* context. The primary research question guiding this thesis is: **What is the emerging geography of the blockchain industry?** To elaborate on this question more fully, the thesis explores three sub-questions:

1. Does the blockchain industry conform to our a priori expectations of the characteristics of new technology industries?

2. How is the industry spread out globally and what are the key geographies of the industry? and

3. How is VC investment into the industry spread out across the globe, and do the patterns overlap with those found in geography of investment and finance literature?

To answer these questions, the thesis draws on a database of 3,839 blockchain firms founded between 2010 and 2018 identified using Crunchbase, an accepted source for analyzing entrepreneurial firm dynamics (Block and Sandner, 2009; Friedlmaier et al., 2018). The data include information on employment, location, and specializations. For firms receiving investments, information on investment rounds was also collected. While some data and metrics were derived directly from the initial Crunchbase dataset, algorithms were developed to enable further detailed analysis and categorization[3].

Elasticsearch, a distributed search and analytical engine was also instrumental in multiple areas of this analysis. Specifically, the ability to create real time data transformations and visualizations on an ad-hoc basis, in addition to analysis-enhancing features such as Significant Terms Analysis made this system an excellent choice for this work. While there have been few efforts to use Elasticsearch in social science literature (see Kononenko (2014), Shah et al. (2018) for exceptions), the engine lends itself well to the needs of economic geographers, ranging from an easy-to-use visualization platform to a sophisticated analysis toolset. This, combined with programming and automation within languages such as Python, enables detailed and sophisticated analysis. Apart from aggregation and trend analysis, the dataset was geocoded, and network data was created at the city-region level to capture investment interactions. Quantitative measures such as Eigenvector centrality were used to identify the key players in the network and examine the network as a whole

[3] Specific scripts will be made available at github.com/mholicka/EmergingBlockchain.

Overall, the thesis adds to current scholarship by examining the firm and industry dynamics of the global blockchain industry. It finds that the blockchain industry appears to conform to understood patterns regarding emerging tech industries. Firms in the industry are small, with most firms having fewer than 10 employees and their application areas are primarily focused on high demand, low capital cost area such as commerce and shopping. Geographically, the industry is global, yet concentrated primarily within world cities, such as Silicon Valley, London, Singapore, and Beijing. Additionally, there appears to be a prominence of second tier world cities such as Tallinn, Estonia and Zug, Switzerland. Investments are primarily concentrated within key locations and the analysis shows that Silicon Valley is the dominant location for blockchain investment and activity, holding a prime position in investment networks. Five locations (Silicon Valley, New York, Singapore, London, and Beijing), who are also extremely well-connected account for the majority of investments. Lastly, supra-regional networks have been observed in the United States and Asia. In this way, the thesis contributes an analysis of the geography of an emerging technology and aids scholars of economic geography in understanding the dynamics of emerging technologies.

The remainder of this thesis proceeds as follows: The second chapter critically examines the literature surrounding technology, geographies of innovation and the geographies of investments. Chapter 3 offers a detailed description of the data and methods used to explore the geography of firm location and investment in the blockchain industry. Chapter 4 examines the results and discuss the potential explanations for the findings, focusing on the three sub-questions identified above, addressing the main characteristics and nature of the blockchain industry, the global geography of blockchain firms, and the global patterns of investment associated with the blockchain industry. Finally, Chapter 5 summarizes the findings based on the analysis of the industry, offers conclusions, identifies the main contributions of the thesis and its implications both for policymakers and scholars interested in clusters, cities, and innovation. In addition, Chapter 5 also identifies limitations and areas of future research.

Chapter 2
Literature Review

This thesis focuses on the emerging geography of the blockchain industry. Blockchain is a decentralized transaction and data management technology with its initial applications being Bitcoin, a cryptocurrency. Blockchain has been attractive to firms and policymakers due to the promise of security, anonymity, and data integrity within the system. Furthermore, it is impossible for third-party control within the system, as only parties operating within the system are able to control it. Each node (party) in the system is able to add information to the system but information cannot be added without all the other nodes in the system agreeing that the information is correct, commonly achieved technically via a proof-of-work algorithm. Most of the current use cases of blockchain related to cryptocurrency, but there are a number of firms that are utilizing the technology in other application areas due to its transparency and traceability characteristics. The main advantages of utilizing the blockchain is the inability to modify or delete data that has been put into the chain. This enables multiple overarching use cases, such as verification, supply chain processing, and data transfer (Cai & Zhu, 2016). With applications within multiple industries, an examination of the geography of the blockchain industry will allow scholars to identify where this technology is taking hold and compare these patterns to widely understood patterns regarding technology clustering and activity. To understand the emerging geography of the blockchain industry, this thesis draws on three bodies of literature: Entrepreneurial ecosystems, cluster dynamics and venture capital (VC) and investment dynamics.

2.1 Entrepreneurial Ecosystems and Innovation

Geographers have extensively studied entrepreneurial firm dynamics as a way to understand regional competitiveness (Spigel, 2020; Spigel & Stam, 2018; Spigel & Vinodrai, 2020; Stam, 2015). This scholarship points to specific factors that help define firm dynamics. These factors can be broken down into three areas: the role of entrepreneurs in tech firms, VC interaction determinants of firm entry. This section will examine these factors in relation to the blockchain industry.

In start-ups and new ventures, the role of the founder should not be understated. In order to start a firm and develop a product, the founder must be intelligent and resourceful. This pattern should hold true in most industries, but the difference between tech and other industries is the pace of innovation and change. Additionally, the newness of both the market, combined with the technology will lead to uncertainty and ambiguity (Delmar & Shane, 2002). Nonetheless, these potential pitfalls are often seen as risks, as the potential rewards for being one of the first and most respected in the industry has major advantages, and entrepreneurs often aspire to become the next Google or Apple of a nascent industry. However, the breakneck speed of tech means that technologies emerge and become obsolescent quickly (Liao & Welsch, 2008). This in turn forces those in the space to become extremely adept at recognizing opportunities and acting with agility and speed to capitalize on them.

Indeed, tech entrepreneurs are found to engage in more start-up activity than non-tech industries (Liao & Welsch, 2008). Firm formation is described in Katz and Gardner (1988) as based on four activities: intentionality, resources, boundary (also known as legitimacy) and exchange. Intentionality has been briefly mentioned above, in which the founder has plans for the success of the business by doing as much as they can in order to be successful. In this step, much planning is done in order to secure talent and finance. Many tech founders have engineering degrees, and as such, they are able to better understand the feasibility of their product, as well as have a higher chance to find the resources needed to get their product to market.

Katz and Garner (1988) additionally find that resource gathering by tech start-ups is a crucial step in the venture creation process. The resources needed to succeed in the advanced tech space encompass both tangible and intangible assets and are different for each industry. The required tangible assets differ from industry to industry, even in the tech space, but the one commonality shared by these startups is the need for intangible assets. Intangible assets come in the form of technologies, associated know-how, and knowledge. Developing and acquiring these tacit resources is more time consuming and difficult than for tangible assets. It is not hard to purchase a laptop but acquiring the knowledge to program can be extremely time-consuming and challenging. In most cases within tech entrepreneurship, there is a level of explicit resources needed, mainly dealing with physical interfaces (such as servers and laptops), but it is the ability to utilize the tech and these physical assets optimally that is the key to survival in tech ventures.

Thus, with the entrepreneur having intentionality and resources, legitimacy and exchange are the last two factor of venture creation as described by Katz and Gartner (1988). Legitimacy of the industry is associated with more firms being created and the technology being developed into a more reliable and stable platform, which sees industrial uses across industries. These firms then participate in exchanges, which has been described in Tornikoski and Renko (2014) as the first sale made by the firm. When the sale is made, the firm is considered to be engaging in “exchanges”.

In previous studies, innovation has been examined at the firm level, with characteristics such as firm size and age highlighted as important. While some studies examined the relationship between firm size and innovation output, there is a lack of empirical evidence to confirm the nature of this relationship (Naz et al., 2015). Furthermore, this is a contested notion. Acs & Audretch (1987) find that firm size has an impact on innovation, but does differ across industries. Mairesse and Mohnen (2001) build on this notion and conclude that firm size does have a positive effect on the innovation in R&D intensive sectors. This conclusion is reached similarly in more recent papers (Baumann and Kritikos, 2016; Czarnitzki and Binz, 2008). However, other studies do not back up these claims. For example, Crepon et al. (1998) find that firm size is not a factor once R&D differences between industries are considered. Similarly, Hansen (1992) looks at firm size and age but find that there a negative correlation of innovation output. Recent scholarship regarding knowledge intensive services (KIS) finds that firm size and age play a role in innovation output. Audretsch et al. (2018) find that smaller, newer firms (referred to as “micro firms”) have a better ability to turn innovation inputs into knowledge as opposed to mature firms. They further identify the importance of a highly skilled labor pool within these industries.

The technology behind blockchain has already seen a plethora of uses in various industries, from the traditional finance application in the cryptocurrency industry, logistics and supply chain tracing, to voting systems. The majority of the newly founded blockchain firms are developing products in application areas that are both in high demand and do not require advanced infrastructure. One key reason for widespread interest in blockchain is that startups within the industry do not require large amounts of tangible assets in order to engage in product development. In the early stages of the firm, where the product is mainly a proof of concept, development can be extremely cheap on the tangible side, but extremely expensive on the intangible side. For example, there are many pre-existing platforms such as Ethereum and NEO with which there are relatively low costs to develop (see (Wu et al., 2019) for a technical breakdown), but development is mainly gated by the knowledge of the programming language and architecture. Of note is that this relative ease of development did not start until about 2015. Before that time, firms needed to start their own blockchain, which requires more tangible assets such as servers, but once again, know-how remains a key factor.

More industrial uses, such as those in natural resources, health, government and military applications, are typically associated with much higher startup costs, due to needing more infrastructure and the complexity of the work. In this way, investments in these areas are expected to be considered niche (Joshi, 2018). Additionally, firms that service these sectors are typically pre-established and new entrants are rare. As a result, new firms will have an extremely hard time entering these industries and firms will often shift their application area to ensure survival. In this case, even if a firm wanted to develop for a more industrial application, unless they have the needed networks, they will be subject to the influencing factor of survival, and will likely switch to an application area that will gain them the investment they need (Sutton, 2000).

Another key to the success of a firm is its ability to create the product. Should the factors align with understood patterns as explained above, yet if there is insufficient capital to innovate, the firm will fail. In order to rectify this, outside investment in the form of venture capital is needed to ensure capital to a firm. An investor provides capital to the firm, in exchange for some financial gain, typically referred as Return on Investment (ROI). In addition to providing capital to the firm, an investor is an invaluable asset to the firm, in the form of an access point to the knowledge network associated with that investor or the syndicate to which they belong (Kang et al., 2020). Indeed, Hellmann and Puri (1999) find that there is a significant reduction in the time-to-market of a product with VC involvement and overall find that relations with an investor increase the competitiveness of a firm (Hellmann & Puri, 1999). Additionally, work done by Chemmanur et al. (2011), which examines the role of VC investment and the efficiency [noted as TFP (total factor productivity)] of firms, finds that overall firm efficiency increases with VC interactions, primarily with improvement in sales. They note that VC involvement has a positive effect on the probability of a successful exit of a firm (Chemmanur et al., 2011).

Literature surrounding entrepreneurial ecosystem and innovation within the blockchain industry points to a number of factors that have an effect on firms within the industry. Using these factors as a baseline, in-depth scholarship paints a picture of low-employee count firms being founded in lower startup cost, high demand application areas. The speed of innovation will cause a rapid identification of market gaps by the entrepreneur, who will find a firm quickly. A small yet highly skilled workforce will be needed, which will develop on top on pre-existing frameworks. Venture capital investment is expected to have a significant impact on the performance of the firm. Additionally, there is an expectation that the location will play a large part in the success of the firms.

2.2 Clusters and The Geography of Innovation

The notion of clusters is not new to economic geographers. Cluster theory was popularized by Porter (1998), who defines clusters as a "geographically proximate group of interconnected companies and associated institutions in a particular field, linked by commonalities and complementarities" (Porter, 1998, p. 199). Further studies have attempted to add to or re-define the definition (see Lu et al., 2018 for a review). There appears to be an agreement among scholars that clustering has a "clear positive impact at the industry level" (Spencer et al., 2010, p. 712), due to a number of positive externalities, such as financial and human capital, knowledge transfer and access to global pipelines. This concentration of firms and/or activity within close geographic proximity has been shown to increase innovation.

Multiple scholars have linked the presence of knowledge spillover to one of the reasons for firm co-location (Leppälä, 2016; Moretti, 2004; Spencer et al., 2010; Sternberg & Arndt, 2001). By co-locating in a given region, firms are able to tap into a deep and specialized labor market, benefit from the spillover of knowledge created by the interaction of workers, and have access to global pipelines (Bathelt et al., 2004a). However, there are some scholars that agree that co-location has a positive relationship with firm and regional innovation, but do not necessarily agree that the presence of knowledge spillovers is the reason for the co-location of firms. They argue that the spillovers facilitated by close physical proximity of workers and firms has less to do with co-location of the firms (Wan & Liu, 2011). Instead, they argue that it is the presence of skilled workers and the ability to establish transaction-intensive relationships, such as those with customers and suppliers that lead to the co-location.

Multiple case studies identify that co-localization is advantageous to firms. For example, Johansson & Lööf (2008) find that there are significant differences in the innovation of regions. In particular, it is found that firms located in Stockholm tend to innovate more than in other regions (Johansson & Lööf, 2008). This is further backed up in the European context by Naz et al. (2015), who find that innovation in Germany is concentrated in agglomerations. These papers appear to back up the claim that place and institutional support matters when discussing clusters.

Indeed, clustering appear to be a worldwide phenomenon, and has been observed across a wide variety of geographical contexts. Ever since the idea gained popularity through the work of Porter (1998), clusters have been a staple in North American economic policy (Council, 2011) . Numerous works point out the benefits of clustering to firms and regions in North America. In the Canadian context, Spencer et al. (2010) find that clustered industries have both higher incomes and rates of growth as compared to their non-cluster counterparts. These advantages are also observed in the U.S, with studies finding that clustering of industries in the US allows those industries to enjoy higher employment and patenting growth (Delgado et al., 2014). Higher employment is found to have a positive effect on innovation, as there is an increase in efficiency, productivity and thus return on investments (Bresnahan et al., 2004; Saxenian, 1994). The presence of externalities within the clusters such as human capital, knowledge transfer and input-output linkages are found to be the reasons for the success. A key observation about clustering in North America is the ease of market integration as opposed to other regions (Pasquero, 2000; Turkina et al., 2016). This eases human and knowledge mobility, and allows geographic forces to create clusters, rather than a forcing of clusters through policy, such as is the case in Europe.

Clustering has been a key element of economic policy in Europe, with targeted plans for facilitating the clustering of certain industries with high innovation potential, such as advanced materials and ICT. Indeed, similar to other regions, clusters in Europe are found to be much higher performing than their non-clustered counterparts (Ketels & Protsiv, 2021). One major difference between clusters in Europe compared to elsewhere is that European clusters are, with a tendency for each nation to have concentrated activity (*The European Cluster Observatory*, 2009). One major challenge is the presence of the “national bias” in Europe. Regardless of the fact that members of the EU should be acting together, there are national interests that often prevent the most optimal clustering. Each country wishes to be seen as innovative, and thus try to create their own clusters, even though this is not optimal, and creates industry fragmentation (Crescenzi et al., 2007). Additionally, unlike the US, Europe’s innovation policy seems to be more focused on equality, rather than optimization. There is a sense of “forcing” of clustering within nation rather than utilizing the geographical processes and capital mobility as seen in the US. Indeed, cluster cooperation is found to not be systematically supported at the regional or national level (Crescenzi et al., 2007; DG Enterprise and Industry, 2007). Additionally, DG Enterprise and Industry (2007) find that clusters within Europe are viewed as a source of regional and national competitiveness, with each nation wishing to be competitive. The major factor for this is the differences in institutions and cultures across the region; a lack of shared goals, philosophies and values can create barriers to cooperation and competitiveness across national borders in Europe.

In Asia, clustering has been used to spur innovation, and clustering has been supported in multiple ways, chiefly through regional and national policies. In a study of Asian clusters, Intarakumnerd & Vang (2006) find that the promotion of industrial clusters is based on the country’s level of development. For example, China and India are large players on the current tech stage, with China being an extremely important player in the current technology space. There is evidence to suggest clustering is heavily encouraged at the national level. In particular, Hinata (2011) explains that China was investing heavily into the technology sector, with multiple programs focused on the creation of a strong tech service sector to rival Bangalore in India. One of China’s most influential programs is the *Thousand-Hundred-10 Project.* The project has the main goal of establishing 10 cities as bases of service-outsourcing, as well as attracting 100 multi-national firms, on top of 1000 local medium and large companies. Combined with the influx of new technology graduates in China, such a program will enable the creation and growth of clusters in the cities (Ōhara et al., 2011). Other nations within Asia, such as Japan have created ‘intellectual clusters’, which aim at creating “region-based clusters of universities, public R&D institutions and knowledge-intensive core companies” (Intarakumnerd & Vang, 2006, p. 4). The aim is to foster interaction between incumbent organizations in the region (universities, firms, and research organizations) to create technological innovation. In addition to the clustering, Japan's patenting process was refined and led to the direct improvement of Japan's ability to innovate. As Guo (2015) concludes "changes in patent applications have a positive effect on growth of TFP, implying that an increase in patent applications reflects a higher level of innovation in Japan" (Guo, 2015, pp. 601–602).

Within the cluster literature, one of the key elements that is often identified as critical to performance is the presence of anchor firms. The notion of anchor firms in cluster literature is not new. Feldman (2005: 312) describes anchor firms as firms that can "attract skilled labor pools, specialized intermediate industries and provide knowledge spillovers that benefit new technology intensive firms in the region". It is often because of these anchors that other firms co-locate with them. In turn, they are able to benefit from many of the same advantages provided by clustering, but already pre-established from the anchor. The anchor will be more likely to provide specialized expertise, physical assets, infrastructure and positive spillover (Teirlinck & Khoshnevis, 2019). Furthermore, firms are attracted to the large labor pool that have the potential to transfer knowledge to the new firms (Niosi & Zhegu, 2005). One of the main arguments for the presence of the anchor firms in a cluster is the growth opportunity and the creation of regional dynamics (Agrawal & Cockburn, 2003). By co-locating with an anchor, smaller firms can tap into the same market and have the opportunity for growth that may not have been present by themselves.

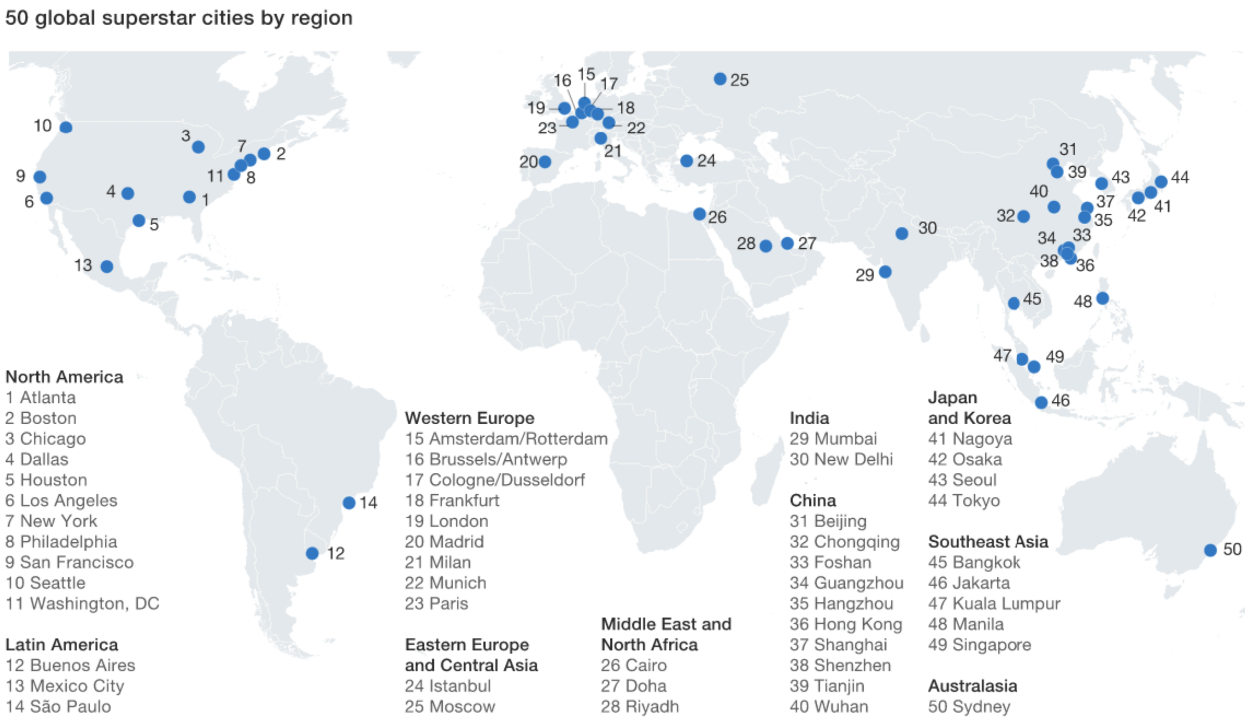
Economic geographers have long established the role of cities in the geography of innovation. The literature agrees that the largest cities provide advantages for innovation (Backman & Lööf, 2015; Moretti, 2004; Wan & Liu, 2011; Wolfe & Conference Board of Canada, 2009). Cities have been found to increase the innovation of incumbent firms due to a multitude of factors, including labor supply, knowledge spillover, and access to global knowledge pipelines. This resonates with a view held in the literature on world cities, which posits that 'world cities' are critical components of the global economy. The idea of world cities is related to the contested notion that there are select cities that hold high levels of concentrated powers and wealth (Smith, 2014). While much literature focuses on a subset of very large cities (e.g. New York, London, and Tokyo) that act as global financial centers (Sassen, 1991), the world cities literature delineates a hierarchical system of cities. Various nomenclatures have been used to classify the position of cities around the world (Beaverstock et al., 1999). What once started off with a select few cities that were both easy to categorize and identify with strong locational identity has now transformed into multiple groupings based on different criteria, evolving into a complex definition. However, the term ‘world cities’ and two related terms - global cities and superstar cities - have been the subject of much debate in the academic community and while they each emphasize different characteristics and elements of cities around the globe, they are often used interchangeably as well.

Contemporary economic geography literature points to the notion of superstar cities, which are based on the “concentration, agglomeration, and knowledge spillovers; government policy impacts; firm-level market concentration; and the all-encompassing need for access to highly specialized pools of talent” (Brail, 2019, p. 5). One such superstar city is San Francisco, which has benefited from a strong regional identity (linked to Silicon Valley) and purposeful actions taken by both the regional and federal government in order to boost innovation and tech, such as relationship building and high levels of investment in key industries to spur innovation. McNeill (2016) describes San Francisco as a city of venture capital supported unicorns. This status attract interest from other firms, due to the benefits associated with clustering and having such a strong infrastructure. Indeed, economic development and locational strategies often base their strategies on access and cultivation of talent (McNeill, 2016). This concentration of talent in firms is also found to increase the innovation of firms (Wolfe & Bramwell, 2016), and it is a key piece in the region’s ability to sustain and increase its economic growth.

The notion of global city was introduced in “Cities of Evolution” by Sir Patrick Geddes in 1915, since then, scholars have been synthesizing the work, and one of the earliest synthesis was Hall (1966) who described global cities as those cities whose are major centers of political power combined with a very strong national government (Hall, 1966). Other factors that were taken into consideration were population, as well as economic power. Further work in the 1970s brought forth globalization into the equation, and its incorporation into the debate meant that leadership was important. With firms spreading worldwide, their headquarter location and the number of headquarters of MNC (Multinational Corporations) in the city would determine its status. Friedmann and Wolff (1982) argue that the role of the global city is to be a control center of the global economy, heavily influencing an economy centered around “economic relations”. Those cities that facilitated the most relations were deemed as global cities (Friedmann & Wolff, 1982). Friedman (1986) goes on to call world cities as “key basing points” for the headquarters of these large and globalized firm, in that all decisions that are needed for the firms go through the main base. At this time, the literature was focused on “hard” power, defined as the built infrastructure (Friedman, 1986). Naturally, this was an easy way to show the power of a city, as the more global the city was, the more money flowed through it, leading to the need for better and better infrastructure. Thus, only the most powerful and thus global cities could afford the best infrastructure. This notion changed slightly in the 1990s, with a split focus on both hard and soft power. Soft power has been defined as strong economic, political and cultural influence (Ilgen, 2013). A global city should therefore hold a strong sense of locational and regional identity that is known around the world.

Despite the use of these three different terms to describe the largest cities in the global economy and their relative position, it is also important to understand the actual location of the cities in question. Figure 2-1 provides a map showing one definition of global superstar cities. The McKinsey Global Institute (2018) uses a pragmatic, data-oriented definition that identifies these cities based on GDP and personal income, identifying superstar cities as cities that have “a substantially greater share of income than peers and is pulling away from those peers over time”(McKinsey Global Institute, 2018). Figure 2-1 reveals that there are global superstar cities on each continent and shows China’s growing prowess; main concentrations occurs in the USA, Western Europe and Asia. Two countries in particular, USA and China dominate, accounting for 42% of the top 50 superstar cities. McKinsey’s study reinforces the theories of scholars in the field. The majority of the top 50 superstar cites are globally integrated, innovative, financial centers. Additionally, 22 cities are regional or national capitals.

Figure 2-1 Global superstar cities[4]



It should be noted that Figure 2-1 offers only one definition; there are many other rankings and classifications systems. For example, Trujillo and Parilla (2016) utilize a combination of factors involving clusters, innovation, talent, and infrastructure connectivity, which classifies cities in seven categories. Beaverstock et al. (2017) utilize concentrations of economic functions, particularly in regard to finances and banking services, as well as marketing, advertising and accounting. Throughout this thesis, the term 'world city' will be used and operationalized using

[4] From The superstar firms, sectors, and cities leading the global economy by Manyika et al., 2018, McKinsey Global Institute (https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Innovation/Superstars%20The%20dynamics%20of%20firms%20sectors%20and%20cities%20leading%20the%20global%20economy/SVGZ-Superstars-Ex5-Expanded-vF.ashx). Copyright 2018 by McKinsey Global Institute

Beaverstock et al.'s (2017) definition of world cities, specifically referring to those cities in the "Alpha"[5] category of world cities.

The importance of clustering and co-location was explored in this section, setting the expectation that the blockchain industry is likely to be a global yet clustered industry, with the majority of firms located within world cities. The presence of positive externalities and advantages brought on by co-location within these cities that have high importance, financial and otherwise, in the global economy is expected to play a large part in firms' location decisions. Particular places, such as Silicon Valley, London, and Beijing, hold status within the world city literature, due in part to strong locational identities, regional policies, and the presence of financial and human capital in higher concentrations compared to other cities.

Additionally, the literature suggests the possibility of a subset of cities, which are up and coming on the world stage due to the presence of anchor firms and/or strong institutional involvement. These cities have the potential to become future hubs of technologies, as anchor firm or institutions provide firms and the regional economy with similar positive externalities and access to the global pipelines as found in clusters located in world cities. Additionally, cities with strong regional specializations or places where regional activity benefits from a specific application are expected to maintain their importance, primarily within the specialized application area. Lastly, one key element of a firm's ability to succeed in addition to location is the ability to gain and maintain capital, typically achieved through receiving investment and financial backing.

2.3 Venture Capital and the Geography of Investment

Investments are one of the ways for a firm to propagate its wealth and influence and diversify its profits to potentially turn their profits into even more profits. The idea is that a successful firm or individual will invest their money into an idea, and thus a firm, with the hope that their money will be utilized to create some service or product that is worth more than the initial

[5] See https://www.lboro.ac.uk/gawc/world2020t.html for an updated list of world cities.

investment or will continue to generate the money back. Investors are typically classified as either angel investors or venture capitalists. The differences between these two groups can be best summarized as varying levels of control on such items as board control, exit requirements and ownership control (Dutta & Folta, 2016). Their study finds that angels and venture capitalists may invest in ways that are complementary to each other, and both are seen as important to startup firms.

Not all investments are successful, and indeed, investment into tech has been found to be very risky in the past, as studies of technology-based firms have reported (Kortum & Lerner, 2000; Westhead & Storey, 1997). However this finding has been contested by Mason & Harrison (2004), which found that there is “no significant differences in the performance of technology and non-technology investments” (Mason & Harrison, 2004, p. 327). They propose three reasons for this conclusion. First, angels do not invest in the initial innovators, rather the followers. Second, the risk of the investments may have been overstated in the papers. Finally, the difference is due to the difference between VC angels and business angels. The findings support Bjørgum & Sørheim (2015), which find that there is differences between technology and non-technology angels. They found a statistically significant difference between the behavior, relationships and their overall involvement (Bjørgum & Sørheim, 2015). More recent articles have suggested that investment is heavily based on the type of firm.

Investments are highly correlated to connections and capital. The more connected an investor is, and the more capital they can provide, the more they can safely diversify their investments and have the highest possible return on investments. Moreover, scholarship has found a strong positive correlation between geographic proximity of investors and their portfolio in terms of success Kang et al. (2020). Success in this case refers to the number of exits and overall funding, which was found to be higher with investors and firms within close geographical proximity(Kang et al., 2020). Taking this into consideration, the most logical place that investment firms should be located is in large, world cities. World cities provide investors with the necessary inputs for success. Investors require financial capital to fund the portfolio, in addition to the knowledge of who and where to invest, or where is a new area of investment. First, investment firms can co-locate with the main institutions, and they are both able to learn from the institutions, as well as develop the relationships needed for a successful relationship. Similarly, being in a capital city means that the global pipelines that have been curated and are associated with the city can be utilized by the investor to extract the best information.

The main take away of the literature is that investors are located in “areas where innovation is high and in high money areas” (Mingo et al., 2018, p. 95). Notably, while there has been some literature regarding the *branching* of the investor offices outside of world cities in an effort to be close to their investments. Investments done by the central office have been found to outperform those of the branches. While the branch may be close to the investment, the advantages of the capital city cannot be understated. Additionally, Chen et al. (2010) finds “that a one standard deviation increase in the number of venture capital offices in a region is associated with an increase in venture capital investments in that area of 49.7%” (Chen et al., 2010, p. 90) Thus, as the city and the investor space within it grow, a positive feedback loop is established wherein investors are likely to invest locally, which prompts further investors and the city grows in power and capital due to this.

This discussion has signaled the fact that there is an expectation that the investor be located in world cities, brought on by the access to needed capital and networks. Thus far, there has been no mention of key geographies, rather a broad analysis. There is scholarship that suggests that these geographies are indeed well known, and points specifically to certain city regions. Nationally, venture capitalist investment within the US has saw a decrease since the 1990s, yet San Francisco and the Bay area [referred to as “Silicon Valley” in this work] remains the “world’s dominant location for startup activity” (Florida & Hathaway, 2018, p. 44). This locale was found to account for 1/5th of global investment. This speaks to the healthy amount of both firms and investors being co-located within this city region, alongside the movement of both angel and venture capital. Global investment overall is highly clustered in specific geographies, with the top geographies accounting for over 50% of total investment. This work additionally contributes to the notion of changing typology within global cities by introducing not just secondary and tertiary cites as Brail (2019), yet an entire tiered subset, noting that there is a difference between emerging and established global hubs.

2.4 Summary

Given this discussion, we can expect that the blockchain industry will conform to well-known patterns associated with the geography of innovation. In this chapter, these issues were explored through three main themes: entrepreneurial ecosystems, geography of innovation and the geography of investments. Each of these sections has provided both an expectation of the results, and theoretical framework to understand potential explanations. The entrepreneurial ecosystem involves relatively new firms with a small, yet highly skilled workforce, with application areas centered around high demand, low startup cost area, typically with a pre-existing technical framework. Investor relationships are expected to be key, and investor activity is expected to be niche within industrial applications. The geography of innovation literature suggests that understood patterns of firm location will likely continue, with a focus on co-location and clustering in world cities. Given this, it is expected that the blockchain industry will establish itself within these world cities. Additionally, a review of the literature suggests that there will be a subset of cities that will be important to the geography of the blockchain industry, and these cities are expected to be supported via a combination of the presence of an anchor firm and high institutional involvement. The examination of the VC literature creates an expectation of both domestic and global investment, with a concentration in world cities due to their connected nature and ability to provide access to capital. In particular, Silicon Valley is expected to continue to be of crucial importance to the industry, as a key location for both investors and blockchain firms.

Chapter 3
Data and Methods

The main goals of this thesis are to understand the blockchain industry through a geographical lens focused on location patterns and investment interaction networks. There are two major data groups that are needed for the analysis presented in this work. Firms comprise most of the dataset, and these are organizations that have self - identified as being a company for their primary role. Rarely is it the case that a firm has access to the capital needed to sustain itself, and this is where the investor dataset comes in. These are firms who have investments into the firms within the industry. These datasets combine both quantitative and qualitative aspects of the firms and investors. Quantitative data such as geographical location, employee count, and number of funding rounds is paired with qualitative data such as firm descriptions, and firm categories. This thesis utilizes both data types within the analysis.

The majority of the analysis in this work will stem from calculations and aggregations performed on one or more of these data fields. In addition to quantitative metric data, geographical data will be utilized to perform analysis at the geographical level, with a focus on the city and country level. Most results in this work will be presented in the form of charts, tables, maps, and figures which have been calculated via these aggregations and put into visual formats where appropriate. Keeping with the fact that is work aims to understand and quantify the relationships between investors and firms on a geographical level, graphing analysis will be utilized to quantify and visualize these interactions.

The rest of this chapter is organized as follows: A short introduction to the dataset and source will be followed by a generalized methodology which will include explanations of some of the algorithms utilized in this analysis. This methodology will serve to explain the data gathering process, cleanup and loading of the data into the aggregation engine, analysis techniques, and exploratory network analysis. Lastly, data limitations will be explored. In order to begin analysis, it is imperative to understand the data source.

3.1 Dataset Sourcing

Data was acquired from Crunchbase, the largest public database with firm and investor profiles. Crunchbase acquires data via collaboration with multiple actors, mainly venture community networks such as Venture Program; community involvement; automated news collection via AI and machine learning; and human involvement.

Crunchbase has been used to some extent in academic world studies. Liang and Yuan (2016) were among the first to utilize Crunchbase in their research, which focused on utilizing the connections among firms and investors in order to predict the relationships. Using the data and statistical models, they concluded that investors are more likely to invest into firms that they share social relationships with, either in a direct or indirect sense. Tarasconi and Menon (2017) utilized Crunchbase to match companies and individuals with patents. They note that Crunchbase has a large selection of micro data and combined with an international dimension, is a suitable data source for database creation. Additional validity for Crunchbase as a data source can be seen in Block and Sandner (2009), in which they utilize the database in order to establish the effect of an economic crisis on venture capital investment. They emphasize the accuracy of the data, finding that the database contains 97% of the interest data, with a Pearson correlation of $r=0.67$ ($p < 0.05$) to National Venture Capital Association (NVCA) data through the timeframe of Q1 2007 to Q1 2009 (Block & Sandner, 2009). They do note that Crunchbase contains a strong focus on the US market. This is one of many limitations with this dataset, which will be discussed in further detail at the end of the chapter.

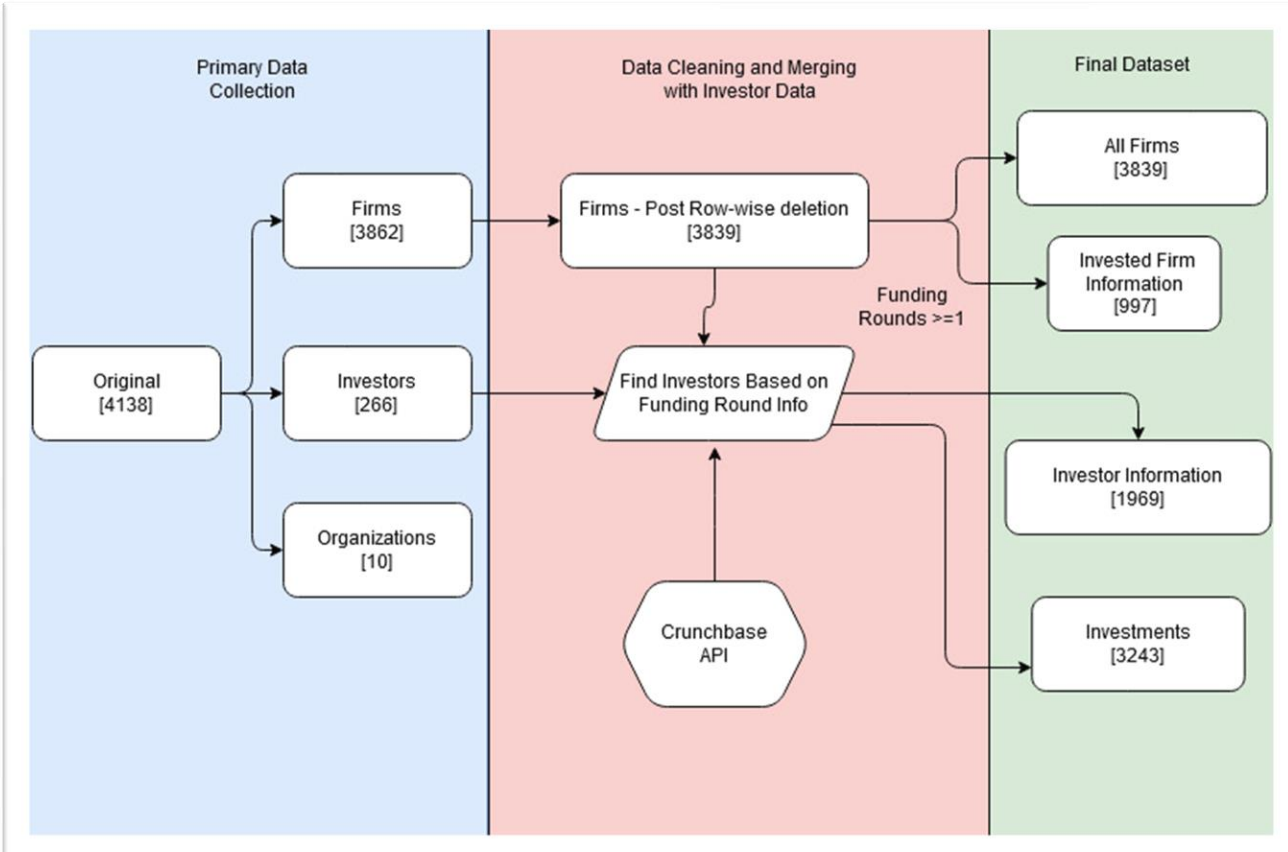
3.2 Methods Overview

The dataset was transformed through multiple steps to create a database of blockchain - related economic activity. The data went through 3 major steps in the process, each with multiple sub-steps. First was Database Acquisition and Cleaning to build a database of investment firms, individuals and blockchain firms, adding in geo-location using available data. Second, this database was loaded into Elasticsearch, a distributed search and analysis engine, for further analysis and visualization utilizing data mappings, and custom analyzers. This allowed for deep analysis of patterns that may not have appeared simply within the data itself. Due to the nature of the methods used in the analysis, data was manipulated and moved from Elasticsearch as needed, with the most common usage being aggregation. With all the data loaded, Data Analysis could begin, in which a combination of traditional statistics, mapping and network analysis were used to create and display the results.

3.3 Database Acquisition and Cleaning

In order to begin any analysis, the raw data needed to be acquired, cleaned, and prepared for loading. This major step consisted of creating the database with Setup, Data Gathering, Cleaning and Geo-Locating. The primary goal of this step was to transform the original data into separate information about the firms, investors, and the investments. A visual representation of this step can be seen in Figure 3-1.

Figure 3-1 Database Acquisition and Cleaning methodology overview



Source: Author

As can be seen in Figure 3-1, the creation of the datasets was split into three separate stages, with the output of the previous step being used as the input of the next step. When needed, additional data was queried from a snapshot of the Crunchbase API, which will be explained in the next section. The two major steps in this stage of the methodology are: primary data collection and data cleaning and merging. These two stages yield the final database.

The first task was to extract the most pertinent information from the raw data. In order to minimize the possibility of changing data, which could lead to changes in the analysis, a snapshot of the Crunchbase dataset was taken, from which all data would be extracted. In this way, current information would be unable to change, and the analysis could proceed without needing to worry about changing information.

The dataset contained all the firms listed on Crunchbase, which amounted to 804,593 at the time of downloading. The interest category in this thesis was limited to blockchain. As such, a filter was used to filter out the rows that contained the category of “Blockchain”. In total, the dataset contained 4,138 rows of blockchain firms. Further splitting of the data needed to be done due to the identification of three major roles and the recorded “primary role” of the rows. While each entry could self-identify with multiple roles, such as investor, company, and school, their primary role was limited to one. This primary role column was the basis on which the rows were split according to their identified primary role.

Table 3-1 Data Overview – Primary role of organization

Type of organization	**Number of Rows**
Company	3862
Investor	266
School	10
Total	**4138**

Source: Crunchbase, Author’s Calculations

As can be seen in Table 3-1, it is not very surprising that the most frequently identified primary role was that of Company, with only a small subset of rows identifying as Investor, and a minimal portion of Schools.

Considering the self-reported nature of the data, there is a potential for exaggeration or misrepresentation. For example, firms may choose to report additional related categories to their main categories in an effort to be found by investors. Crunchbase utilizes both manual and automated analysis of these self-reported fields, with the aim of establishing data validity.

As this paper is mainly focused on the blockchain firms and their investors, the dataset was limited to the Company and Investor roles. It is clear that the investor role does not represent the full set of investors, as these are only the investors that identify as being in the blockchain category, which does not account for venture capital firms. Firm investments were split into funding rounds, which contained the data regarding the investor and - if available - financial contributions. Utilizing the unique ID of the investor within the funding round data, investor information was acquired. Individual investors and their investments were identified and linked to the companies. For example, if Firm X was found to have a funding round of unique Identifier Y, with investors A and B, these unique identifiers could then be used to get the needed information for A and B.

Investor information was added, increasing the number of investment firms from 266 to 1969. This was not complete however, as individual investors were also present in the funding rounds, so adding their information was necessary as well. A summary of the investment rows can be found in Table 3-2.

Table 3-2 Data overview - investor type

Type	Number of Rows
Individual	589
Organization	1969
Total	**2558**

Source: Crunchbase, Author's Calculations

Table 3-2 demonstrates that while organizational investment is prevalent, individual investment is also present. At this point, the database consists of the investments, the investors, and the firms. This data can be thought of quite raw at this point, with multiple potential missing data fields, incomplete geographic information and more. The next step was to clean this raw data, and geolocate it.

As this paper is interested in both spatial and aspatial trends of blockchain investment, and with a lack of geo-coding provided from the base data, it was necessary to perform the geo-coding of the rows. In this process, each row was assigned a latitude and longitude based on a hierarchal scheme of the locational data. At the same time, data integrity challenges such as incomplete data were addressed. The main challenge that was faced in terms of data integrity was missing data, mainly firm location. A possible solution to this challenge was to utilize row-wise deletion of data that did not contain locational data. Row-wise deletion is commonly used in firm-level studies; for example, exclude firms with errors in the data. Other substitution methods such as pairwise or mean, or imputation methods, as described by works such as Graham, 2009;Kang, 2013 are impossible in this case, as there are no additional data to substitute or impute. This resulted in an exclusion of 23 out of 3862 firms and 8 out of 1969 investors due to a lack of location data. Similar deletion was done in the case of categorical analysis if the "category_groups_list" field was missing. This resulted in the exclusion of 284 of out 3862 firms and 353 out of 3243 investments. The deletion of this data risks removing some potentially key interactions in the case of investor firm relationships, as well as removing some nuance and potential region specialization due to the exclusion of firms that did not have information about their activities by category. Imputation methods such as description analysis or the utilization of the category_list field were found to be not possible due to the presence of only "Blockchain" as the category, which does not reveal a more detailed application area. Additionally, due a difference in the data characteristics associated with the investment organizations and individuals (such as a first and last name, gender, and featured company), these two groups were treated separately throughout this entire step.

With data integrity checked, geo-location could occur. Geo-coding is the act of turning text to longitude and latitude coordinates. Examples of geo-coding is turning an address to a specified longitude and latitude. For the purposes of this study, geolocation was done in a multi-tiered fashion, with more detail (large scale) being preferable to broad (small scale) geocoding. This allowed for the substitution of missing data wherever possible and ensured that all relevant data was kept. In this case, the data available for the rows – in order of preference was postal code, address, city, region, state code, country code. In most cases, this hierarchy was followed, and it was common that if postal code or address was missing, so was the other, leading to the city being used. This city-based approach was sound for the dataset at the time but proved to be lacking in the nuance upon further examination, with Silicon Valley being the main cause of bias.

Through initial dataset examination and exploratory analysis, it was clear that cities within the Silicon Valley region posed a potential bias in the analysis with their large firm counts in proportion to other cities. To minimize the effect of an over-representation of these cities in the analysis, the super region of "Silicon Valley" was created, which aggregated all cities in Silicon Valley as defined by the Silicon Valley Historical Association (*Where Is Silicon Valley?*, n.d.)[6]. Notably, this approach is subject to potential author bias, and the reasoning behind this choice is described in further detail in Section 3.6

Given the well-known importance of Silicon Valley in the global technology industry, this adjustment allowed for the better capture of the blockchain activity in the region. Additional considerations for city regions such as New York (e.g., Brooklyn, Manhattan) were implemented in a similar fashion. Throughout the course of analysis, no other city region was found to be overrepresented in a similar way, and as such, the remaining regions were not modified. With all the data geo-located, data cleaning and geo-location was completed.

The main goals of Database Acquisition and Cleaning were to gather the data, clean it and geo-locate the companies. At the end of this step, the data was separated into four distinct tables: blockchain firms, investment firms, investment individuals and investments. Each of the data tables with firm or individual information were geo-coded. In order to better visualize and analyze the data, Elasticsearch was chosen as the primary data storage and analysis tool.

[6] The full list of aggregated cities can be found in Appendix A.

3.4 Elasticsearch

The large amount of data required for the analysis of this work required a data solution that was flexible, customizable, and allowed quick analysis. Elasticsearch (ES), a distributed search and analytical engine was chosen to store and analyze the data. The main advantage of Elasticsearch is the ability to index all data types, from text to geographical coordinates. This allows for real time data transformations, queries, and visualizations to be made. The main contributing factor to the usage of Elasticsearch was the relative ease of creating complex aggregations throughout the analysis. Despite the advantages of a system such as this, there have been few applications of Elasticsearch within the academic community (see Kononenko (2014), Shah et al. (2018) for exceptions). In order to better understand methodologies involving ES, a list of a list of terms and their RDBMS [Relational Database Management System] counterparts are shown below (see Table 3-3).

Table 3-3 Elasticsearch vs RDBMS Terms

Elasticsearch	RDBMS
Index	Database
Mapping	Table
Document	Tuple

[modified from A framework for social media data analytics using Elasticsearch and Kibana (Shah et al., 2018)]

One of the key advantages of ES is that the user is able to create the mapping as they are needed, and dependent on the document type. This allows for both automated and custom analyzers to be used. For example, a *keyword* document type allows for the searching and filtering on that document (for example searching by *city*), which is very useful for quick and efficient filtering and searching(Elasticsearch, 2021). Similarly, *text* fields such as descriptions are mapped and may include custom analyzers which aid in extracting the core of the text. The usage of custom analyzers for text was used to separate categories from each other, allowing for deeper analysis of the categories and allowing for categorization. The main custom analyzer for this work was used to filter all the stop words and separate on commas. This allowed text such as *"Data and Analytics, Financial Services, Information Technology, Privacy and Security"* to become *"data analytics", "financial services"," information technology", "privacy security"*. This splitting allows for much greater analysis of the categories of firms in the same space. Before, the only firms that would be aggregated together in a city would have to have all *N* of the categories in the list. With the custom analyzer, they only need to have 1 of the *N* categories to be aggregated. For example, without the analyzer, if Firm A has categories "*Data and Analytics, Financial Services*" and Firm B has categories *"Data and Analytics, Information Technology",* they would not be aggregated by categories, leading to incomplete analysis.

Keywords and geographic coordinates comprised most of the mappings. This allowed searching, filtering, and aggregation of the terms alongside visualization in the form of maps[7]. The mapping structure was modified throughout the process, as data was added, and fields changed in utility to the work. One of the main reasons for custom analyzers to be used is for the purpose of Significant Terms Analysis.

When there is a large amount of data, connections and trends can be skewed towards the majority, and may not reveal the most interesting connections, or who the key players are. In utilizing unique firm counts by location, an analysis of relative importance may aid in the identification of patterns which may not be visible in the dataset. In order to distill and find the true key players, both from an investment and firm perspective, the "Significant Terms" aggregation was used. In effect, this aggregation utilizes Document Frequency / Inverse Document Frequency (DF/IDF) to establish the relationships that are specific to the query, which is a technique to determine relevance of documents in queries (Jalilifard et al., 2021; Ramos, 2003). In short, the model uses a ratio to determine the rarity of the term within the dataset, compared to the appearance of those items within the result set. If the document appears often in the result set, yet is rare in the overall dataset, it is thought as significant for that search query. For example, there

[7] A full listing of mappings and analyzers can be found at github.com/mholicka/EmergingBlockchain/Mappings.

were 2,222 total investments from US-based investors to other locations globally. A simple top 5 by total investment analysis produces the following countries: USA, Great Britain, Canada, Singapore, and China. Digging deeper, utilizing Significant terms, the 5 most significant connections are USA, Canada, Israel, Cayman Islands, and Egypt. This means that these countries are proportionally seen more often in this particular query than they should be based on the dataset, and as such can be considered of higher relative importance. This could signify a strong and particular relationships between the two countries.

3.5 Data Analysis

Two main forms of data analysis were undertaken: Exploratory Analysis and Categorization, and Network Analysis. Exploratory Analysis aimed to describe the dataset metrics, including distributions, firm metrics, and geographies. Categorization was used to detail the analysis by relating patterns of firm activity and investment to their product category within geographic regions. Lastly, Network Analysis was conducted, to establish firm-level connections on a city-level basis. This step aimed to examine the geography of the industry and establish cities that are integral to the network. In this way, the global players at the city level can be found and described.

3.5.1 Exploratory Analysis and Categorization

Once the index was loaded, exploratory analysis could begin. In this phase, the main goals were to examine firm structure, as well as spatial and temporal relationships between firms. First, Top 20 analysis was used to examine the highest-ranking hits for a given statistic to discern possible trends in the data. Top 20 analysis statistics included: spatial distribution of investors, firms, investments, and temporal distribution of firm founding. Spatial distribution was measured in unique count of firms/ investors. A discussion regarding this decision and the limitations it introduces can be found in Section 3.6. Similarly, firm structure statistics, such as employee count and category involvement, were used to provide additional evidence for trends and highlight key players for a given statistic.

As noted in the introduction to this section, the potential presence of regional and national specializations was unaccounted for in the original dataset. The data did contain two fields that listed specific self-identified categories of the firms, being *category_list* [529 distinct categories] and *category_groups_list* [46 distinct categories]. The decision to utilize the *categories_groups_list* field was made with respect to the fact that categorization could be done with consultation of industry articles, whereas categorization on such a high number of categories was subject to bias and subjectivity. Thus, it was necessary to categorize the firms and investors into respective categories.

As mentioned above, there were 46 distinct categories within the *categories_groups_list* field, and upon further examination, some of these categories could be grouped together, in a super-category. Seventeen separate super-categories were created with consultation to industry articles, and extremely similar categories were categorized to one super-category. This enabled overall categorization to be more accurate as extremely prevalent categories such as "internet services" were not overshadowing "true" categories such as healthcare[8].

Categories with similarities such as "consumer goods" and "clothing and apparel" could be put into the same super category of *commerce and shopping*. In order to categorize the dataset, an algorithm was needed. The goal of the algorithm was to utilize the logic of categorization as described above and automate the categorization of the firms. Upon further examination, the naïve solution, which would simply categorize based on the sum of categories within the list typically yielded either *finance* or *software and analytics*, as these were by far the most represented in the listings. This had the potential to detract from the true category, as most of these firms would be in the software field, yet this did not bring the needed nuance to the categorization. As such, a tier list of categories was created, with the goal of assigning tiers to categories based on their overall prevalence and how much of a firm's activities could be explained by this category. For example, *internet services* was too general to adequately capture a firm's activity profile, yet *real estate*

[8] The full mapping chart can be found in Appendix B.

provided a more specific view of the firm's activities. Table 3-4 describes the tier list and the associated categories.

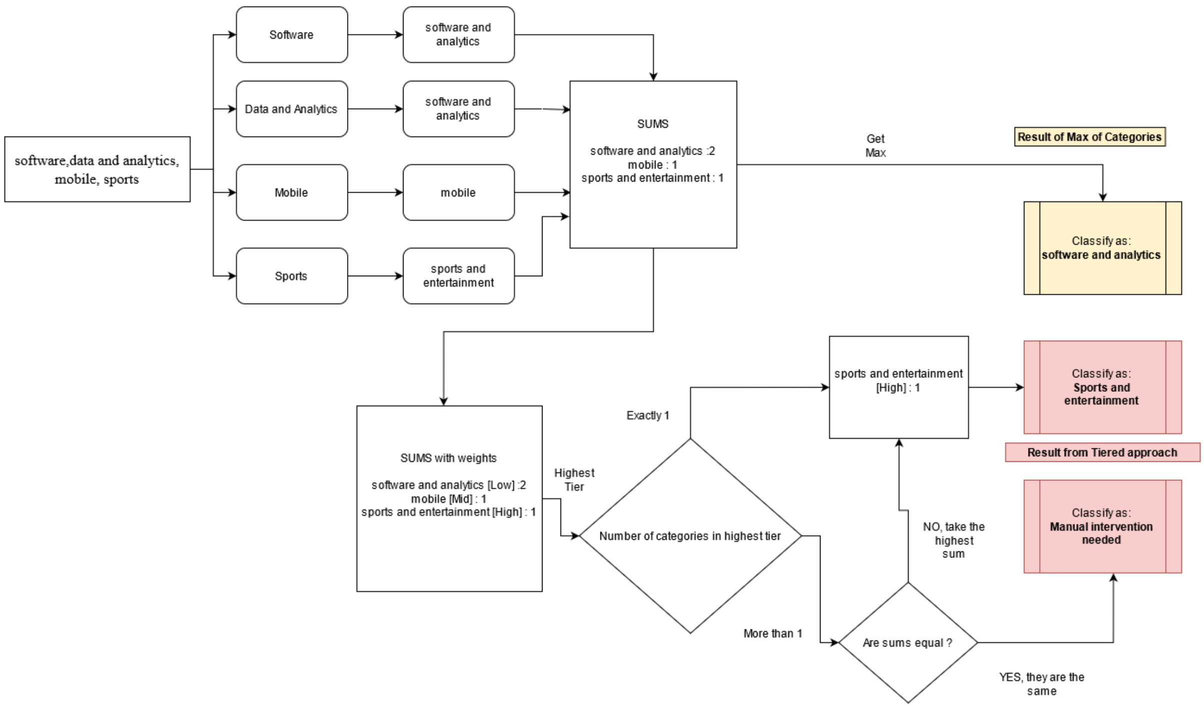
Table 3-4 Tier list of categories

Tier	Categories	Explanation
Top	• Health • Natural resources and Energy • Sports and entertainment • Personnel • Education • Property and real estate • Government and military • Logistics and transportation • Food and agriculture • Community and lifestyle	These categories were found to be the main identifiers of a row. They may be associated with lower tier categories, but the rows that have these categories were found to be best categorized based on that particular super category.
Mid	• Commerce and shopping • Privacy and security • Science and engineering	While some rows identified as just these categories, in most cases there were additional Top tier categories.
Low	• Software and analytics • Finance • Hardware • Mobile	These categories were found to not be the main category. They were usually combined with one or more of the higher tier categories.

Source: Author

In order to better understand the categorization of a firm, an example is helpful. Consider Firm X, which reports to have activities in the following four categories: *software, data and analytics, mobile, sports.* These are assigned to the super categories as follows: software [software and analytics], data and analytics [software and analytics], mobile [mobile], sports [sports and entertainment]. While there are more instances of *software and analytics (2),* the algorithm assigns the firm to sports and entertainment because it is in a higher tier. A visual explanation of this process and a comparison to the naïve maximum of category is summarized in Figure 3-2.

Figure 3-2 Visual categorization methodology



When this specific firm was examined in detail, it was a mobile gaming platform that utilizes blockchain, providing some confirmation to the utility and approach of the algorithm. Additionally, this example serves to reinforce the importance of the tiered approach. Should the tiered approach not be present, this firm would be simply categorized as *software and analytics*, obscuring the actual activities of the firm and potentially contributing to the loss of nuance in the analysis. This algorithm will have challenges when there are no maximums based on the tier list. Consider Firm Y with the super categories of *software and analytics, health, education.* Since education and health are the same tier, manual intervention was needed to identity the category. In this case, the description and website content were examined in order to establish the appropriate category. There were only 31 instances where manual intervention was needed.

The main use case for the categorization of the firms and investors was to examine the spread of the categories and examine the reasoning for potential concentrations of blockchain activities, which may be due to regional specialization or national policy focus (Piras et al., 2012; Sacco, 2017)

3.5.2 Network Analysis

Network analysis has been used in social science research to visualize and analyze connections between entities (Benali & Burlat, 2012; Heemskerk et al., 2016; Fagiolo et al., 2010). In economic literature, agents tend to benefit from connections to others, and those agents better positioned (i.e., more central) are able to benefit more than non-central agents. This is especially common within knowledge-based industries, as it is the constant change and evolution of knowledge, alongside discoveries of new knowledge, which advances industries. Centrally connected agents are able to tap into the overall network and be at the forefront of the knowledge, with all relevant information going through them. They are also typically associated with high levels of human and financial capital (Khan et al., 2019). In this way, these agents can be thought as the most important to the industry. Intuitively then, the expectation is that the most central nodes in the network would be located where the industry has already clustered, or where there is a strong locational identity when it comes to innovation.

In this case, place-based connections between investors and firms were examined at the city-level. The city-level was chosen because this was the most detailed level of analysis possible while maintaining anonymization of the firms. Thus, for this analysis, the agents (nodes) are the cities themselves and the interactions (edges) are the sum of all undirected interactions. For example, there were two investments from Hangzhou into Beijing and 7 investments from Beijing to Hangzhou, thus there are nine interactions between these cities.

Throughout the network analysis, it was imperative to focus on meaningful relationships. For this reason, only relationships between places where there were at least 10 interactions were included. Sensitivity analysis was conducted to ensure that major interactions were captured, and noise was reduced. The threshold of 10 was found through a visual analysis of the network. Beginning with no threshold, the threshold was increased until the interactions on the map were clear from a visual standpoint. Lowering the threshold resulted in a very busy map, in which examination of the interactions would have been challenging. Setting this threshold helped to avoid such methodological pitfalls such as low firm count cities being shown as heavy investment places. The goal of this exploratory network analysis was to find those nodes, or cities that are key to the network. In networking terms, this thesis is mainly concerned with the *centrality* of the node. Social science literature has utilized multiple algorithms to measure the centrality of a network. Common algorithms for this purpose are degree, closeness and eigenvector centrality (Wang & Street, 2015), each of these are discussed below with respect to how they are used in this analysis

One measure of node influence is Eigenvector Centrality. In this algorithm, each node's links are summed in order to get the total *degree* score. The degree of a node in this case is simply the number of different countries which have a relationship with the node. These so-called first-degree connections are important, but further we can dig into these connections and their subsequent connections, the more information that can be gathered about the nodes in the network. Consider a strongly linked country by way of first-order connections. That is, a country that has a large number of connections to other countries. Previously in this paper, the importance of networking and global pipelines was discussed. Considering that innovation and funding typically comes from the main players (in this case the US, UK, and China), having a strong relationship with those countries would enable the country to be even stronger. This is where eigencentrality, through multi-order connection, allows us to see the strong nodes within the network. By examining the degree of the connected nodes in addition to the first-degree connection, the strongest links can be found. Cities with strong eigencentrality will be globally connected and have more direct access to the global pipelines, which enables them to innovate more effectively.

Economic literature has some examples of eigenvector centrality, with some precedent being set by (Kaiser, 2017), in which they utilized this centrality measure to understand the importance of countries in a connected network. Based on the literature review in above sections, there is an expectation of core nodes and periphery nodes in this network, and this notion was found to be the case with the use of eigenvector centrality in this network.

3.6 Limitations

When dealing with business data, especially when taken from one source, a number of potential limitations appear in the data. In the case of the data used in this thesis, there were several limitations. These limitations include source bias, author bias, firm self-reporting and missing/incomplete data. Each of these limitations is discussed below. The initial limitation comes from the fact that the data comes from a firm in the United States and as such, it opens up a possibility for a bias in which there is an overrepresentation of the firms in the United States. While this remains a limitation, numerous works, which were outlined in Section 3.1 have noted the accuracy of the dataset, and as such, this dataset remains valid for the analysis. Similarly, this opens up incentive for firms in the US to have more detailed reporting, contrary to other countries, for which the data may not be full due to a lack of reporting or the unavailability of data.

The dataset has experienced manipulation throughout the course of this analysis. Each intervention and decision by the author subjects the dataset to potential bias. Major areas of potential author bias include the classification of firm by application area and the geocoding. The classification of firms was done with consultation with industry articles on the nature of the application area, and informed decisions were made based on this data. While every effort was made to ensure that there was a seamless transition of data throughout the manipulation, there was some data loss due to a lack of available data. Only those firms missing locational and application area data were discarded, as these were needed fields throughout this analysis.

The dataset for this work is comprised in part with self-reported data. By its very nature, this data may not always reflect real world values. This challenge would be propagated through the analysis, as under or over-reporting of application areas would subject the algorithm to different results. Likewise, an inaccurate reporting of locational data would have the potential to skew the analysis. Understanding this, previous scholarship regarding the accuracy of the data source was consulted (Block & Sandner, 2009). With a positive result regarding accuracy to real-world values and the presence of both automated and human verification throughout the data source, the author saw no reason to distrust this self-reported data.

Furthermore, the data underwent much transformation from the original dataset to the final data. Multiple steps in the process forced deletion of firms' information due to a lack of geographical or categorical data. Each firm removed from the data has the potential to decrease the robustness of the analysis. Data preparation was a key aspect of the analysis, and multiple steps were taken in an effort to minimized data loss. However, there were multiple instances of rows without any locational data, and considering the need for location in the analysis, it was necessary to remove them. One key aspect of analysis of this thesis was understanding the data present and the limitations brought on by utilizing certain data versus others. For example, while the categorical data contained the most information, it was not used in the network analysis portion, in favor of the larger dataset, which did contain the needed geographical data.

In addition to economic metrics, geographical data was utilized extensively throughout this analysis. Similar to the above economic data limitations, geographical data is subject to author bias. Bias in this area may have arisen primarily from the aggregation of city regions and the analysis of spatial distribution. As mentioned earlier in this chapter, the cities within the defined region of Silicon Valley were aggregated in an effort to reduce the presence of primarily cities within this region and enable a more global analysis. Notably, this approach is subject to some challenges. The first being the creation of the aggregated region causing an artificial concentration, and the second being the fact that this aggregation was repeated only for certain known city regions (such as the boroughs of New York), and may not be globally encompassing, thus creating an imbalance within the regions as opposed to others. Silicon Valley is commonly referred to in the academic literature, despite spanning multiple cities (Bresnahan et al., 2004; Breznitz, 2014; Etzkowitz, 2019; Saxenian, 1994); this is also true when dealing with New York (Wolf-Powers, 2005). Additionally, challenges arise from further aggregations globally. Not only is a dataset of aggregations challenging to source from reviewed sources, but there is debate on approaches to spatial aggregation and defining the boundaries of metropolitan regions. For example, some scholars have considered whether cities like Zug and Zurich in Switzerland should be considered part of one metropolitan region (Dessemontet et al., 2010), while other scholars have treated them as separate city-regions (Gugler & Keller, 2009; Kondova, 2018), suggesting the viability of both approaches. Due to the challenge of obtaining a data source that would provide the needed aggregations, aggregation was only done for Silicon Valley and New York. It should be noted that by creating these aggregations, concentrations may be artificially inflated, thus resulting in a skew in the results.

In this analysis, concentrations of firms are assessed using the number of firms and investors. This concentration measure is not ideal and has the potential to overestimate the importance of a particular place compared to relative measures. Consideration was given to using additional measures of concentrations, such as location quotients [LQ] or percentage measures. Unfortunately, these approaches were not pursued due to a number of challenges related to the data. In order to calculate location quotients, data would be required on the firms within the blockchain industry at the regional and global level, as well as for either the universe of firms or the universe of technology firms. In the case of the latter, defining the tech industry is extremely challenging and the subject of debate (Lamb et al., 2016; Wolf & Terrell, 2016). In addition to there being competing definitions of 'tech', these definitions all rely on well-defined statistical classifications of industries, which are absent from the Crunchbase data. Thus, it would be extremely difficult to operationalize this approach given the structure of the Crunchbase dataset. Similarly, the processes involved in the cleanup of the blockchain dataset posed significant challenges and the work to extend this from the small number of blockchain firms to the overall dataset (over 800,000 records) was beyond the scope of the project. Despite these limitations, there is a strong case for using this novel dataset.

Chapter 4
Results and Discussion

This chapter aims to answer the primary research question: What is the emerging geography of the blockchain industry? The analysis is divided into three components. The first section on the nature of the blockchain industry examines the blockchain industry through firm metrics and application area. It examines if the blockchain industry conforms to the understood patterns associated with the broader tech industries. Second, the geography of the blockchain industry is examined in terms of both firm and investor activity and application area. In this way, the key firm concentrations can be identified and places where anchor firms and institutional involvement may be central can be explored. Finally, the geography of investments is examined through the lens of the connections between investors and firms in the blockchain industry. Exploratory network analysis compares the dynamics of investment within the blockchain industry to patterns documented within relevant scholarship. The chapter concludes by offering a summary of findings and possible implications.

4.1 Nature of the Blockchain Industry

Firms in the blockchain industry utilize innovative technologies in their work. To utilize innovative technologies to their full potential, a large amount of human and potentially financial capital may be required. Gaining an understanding of the firms within the blockchain industry may allow for a better understanding of industry dynamics.

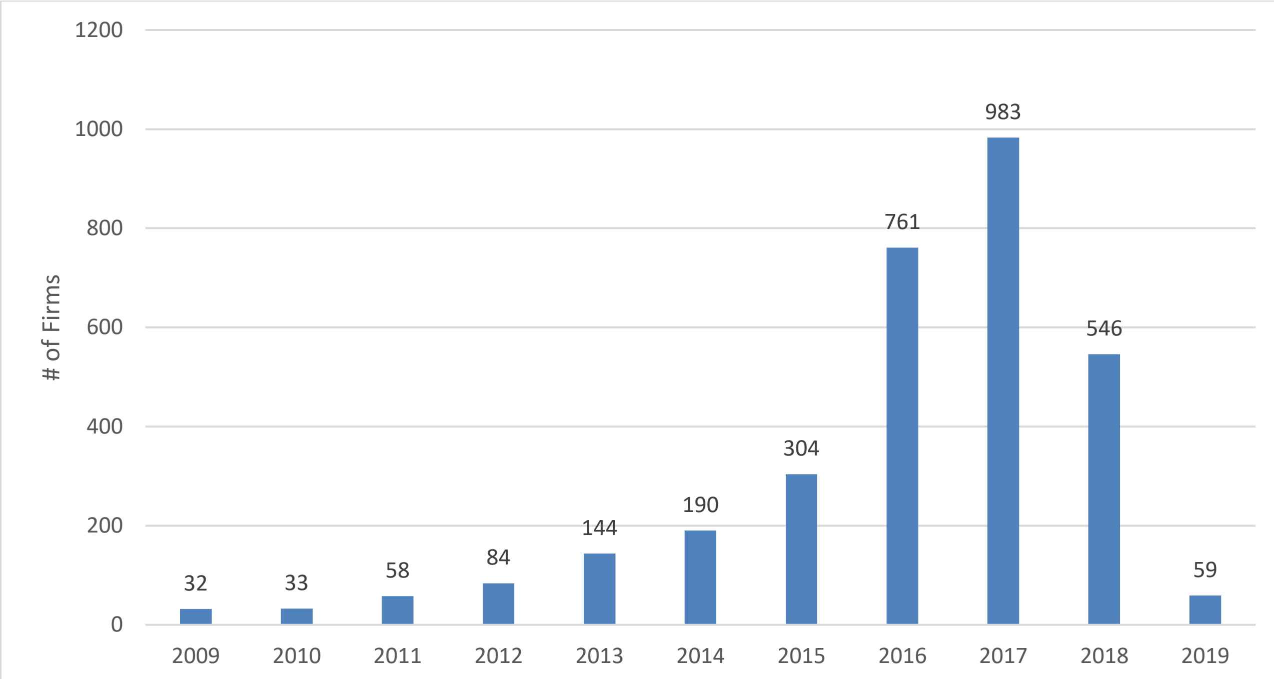
Gaining an understanding of the firm metrics will allow for further analysis of existing trends. This section aims to give the reader an understanding of the blockchain industry in terms of metrics and application area, with the role of the investor being examined in regard to these metrics. In this way, the investment attractiveness of a firm can be examined, and these findings can be explored through a geographical lens. This analysis also examines when firms were founded, in which potential explanations for firm size can be given. Based on the literature review, there is an expectation that firms will be small and there will be a high number of new startups.

Finally, the application areas of the firms are examined. Combined, these metrics allow the reader a window into understanding the blockchain industry. With this background in place, the role of the investor can be examined. Potential patterns of investment can be analyzed and explained. Additionally, the analysis aims to identify the most prevalent investment types in this industry. The chapter concludes with a summary of the findings and endeavors to answer the section's guiding question: *Does the blockchain industry conform to our a priori expectations of the characteristics of new technology industries?*

4.1.1 Firms in the Industry

Figure 4-1 (below) shows when the firms in the dataset were founded. As blockchain technology was only released in January 2009, the chart begins with firm founding in this year. As the chart shows, it took time for the technology to be adopted. Note that the analysis excluded 118 firms founded prior to 2009; these firms are mainly adopters of blockchain rather than specializing in the development of blockchain applications.

Figure 4-1 Firm founding per year



Source: Author's Calculations

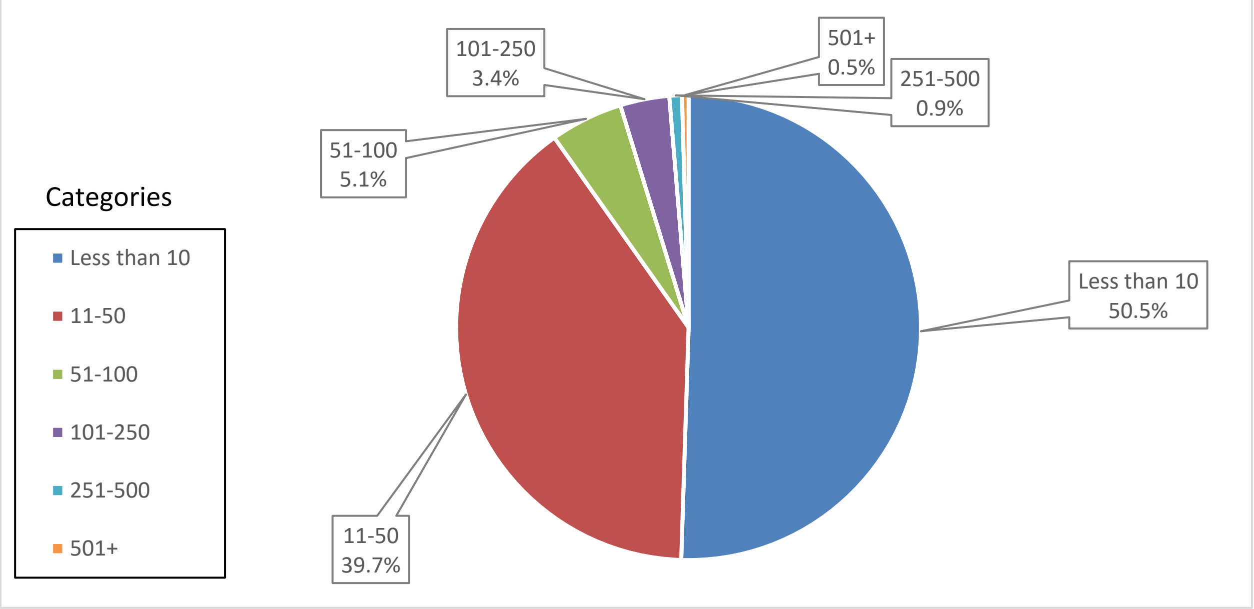
Note: missing data excluded

The data shown in Figure 4-1 demonstrates a gradual increase in the number of blockchain firms founded between 2009 and 2015, followed by a large increase between 2016 and 2018. The two years of growth in firm founding as presented in this dataset were 2016 and 2017, wherein 54.6% of all firm founding's occurred post 2009. This gradual increase in firm founding falls in line with patterns of emerging industries. Throughout the early years of a new technology emerging, startup firms are founded that use the technology in the hopes that they will somehow stand-out from other firms in the field. The industry may be thought of as innovative yet cannot be thought of as a safe enough investment for the majority of the investors.

There were regulatory shifts that changed the landscape of investment in blockchain technologies, potentially explaining the decrease in the number of firms founded. For example, in September 2017, China banned ICOs (Initial Coin Offerings). ICOs allowed firms to crowd-source investment in exchange for financial gain upon success of the firm (Schückes & Gutmann, 2020). This in turn meant that traditional funding methods could be replaced with ICOs in those firms that were unable to get funding. The Chinese ICO ban was found to have regulatory spillover effects that decreased the ICO volumes throughout the world (Bellavitis et al., 2020).

The blockchain industry is comprised primarily of small firms. Figure 4-2 shows the distribution of firms by size.

Figure 4-2 Employee counts



Source: Author's Calculations

Note: missing data excluded

For the most part, blockchain firms are small, with 91% of all firms being between 1-50 employees, as shown in Figure 4-2. There may be several reasons for this. First, blockchain is a new and emerging industry, and as such few firms will have established themselves in the market and scaled up. Second, another possible explanation relates to the low barriers of entry in engaging with blockchain technologies, relative to other technologies. Blockchain as a technology is advanced and challenging but does not require the same amount of physical capital and overhead investment for its utilization as some other advanced technologies. In order to use blockchain, it is more important to have skilled human capital, technical connectivity, and business relationships rather than access to supercomputers, as most blockchain applications can be developed on any laptop if the firm has the pre-requisite knowledge of the frameworks and algorithms. Firm size provides a brief glimpse into the industry and reveals what appears to be a startup culture within the industry.

The application area of a firm can help to describe patterns that are being found. One observation in the literature was the notion that firms in this space will typically align themselves with an application area that can be considered less capital intensive in terms of creation costs, and also an application area where there is market demand. Such applications are typically within marketing and software, as these can be seen to fit the pattern of low technical requirements and barriers to entry, and high demand. Table 4-1 below shows the main applications areas associated with blockchain firms.

Table 4-1 Application Area

Industrial Application	# of Firms	%
Commerce and Shopping	1038	26.9%
Software and Analytics	816	21.1%
Finance	352	9.1%
Sports and Entertainment	271	7.0%
Privacy and Security	246	6.4%
Hardware	242	6.3%
Health	110	2.8%
Science and Engineering	110	2.8%
Logistics and Transportation	93	2.4%
Natural Resources and Energy	67	1.7%
Personnel	60	1.6%
Property and Real Estate	60	1.6%
Education	32	0.8%
Food and Agriculture	25	0.6%
Mobile	23	0.6%
Community and Lifestyle	17	0.4%
Government and Military	16	0.4%
Uncategorized	284	7.4%
Total	**3862**	**100.0%**

Source: Author's Calculations

As can be seen in Table 4-1, while the application area of firms is diverse, a high proportion of applications are focused on services. Of note is the 7% uncategorized area, which is the fifth highest by percentage. The lack of information from these firms may result in a skewed analysis, but the lack of data and the inability to impute this information leaves little choice in the matter. In addition, based on the trends regarding application discussed in previous sections, there is an expectation that this chart will increase throughout the top categories, and no changes in placement would occur. The area which this lack of data could yield more nuance is the lesser categories, such as government and military. Being subject to a low n challenge, the uncategorized firms could add to concentrations to these areas, potentially strengthening the analysis. The top 5 categories, which account for 70.5% of all firms in the dataset, are application areas that follow the trend of low start-up costs, and high demand. For example, firms in commerce and shopping can rely on pre-existing platform or framework such as PAX for loyalty-based cryptocurrency rewards (Bülbül

& İnce, 2018), rather than creating novel software applications. Additionally, the number of firms within this specific category could indicate a proof of market interest for a founder.

Industrial applications such as Health, Logistics and Transportation are not nearly as popular. The reasoning behind this could be that firms in the space are already backed with a specific project or goal in mind by a large player in the industry, or that the demand for this technology in the application area is more limited. It is conceivable that demand for commerce and finance-based application would be bigger than demand for government and military or health applications. We can explore this phenomenon further by investigating the role of the investor.

IPOs (Initial Public Offering) and acquisitions have previously been utilized as a measure of success(Guo & Zhou, 2016), in which firms that reach this status have been shown to be successful in the long run. Only a small percentage of the firms within the dataset [2%] have had an IPO or have been acquired, which indicates that there is a large majority of small, operating startups in the industry, which would conform to the understood patterns within literature. The status of firms in the dataset is found in Table 4-2.

Table 4-2 Firm status

Company Status	Number	%
Operating	3762	97.9
Acquired	50	1.3
IPO	27	0.7
Total	**3839**	

Source: Author's Calculations

Table 4-2 paints a picture of an emerging industry with only a few success stories. Most of the firms in the dataset are operating, but only a few have gone onto an IPO, or have been acquired. What this could indicate is that only a small percent of firms have demonstrated the effectiveness of the product, yet they are an example of success, and other firms will strive for that success. Further examination into these dynamics is beyond the scope of the current inquiry.

4.1.2 Firms Receiving Investment Funds

A possible measure of success for firms is the amount of outside investment received. This allows for the firm to gain access to the network of their investor, which in turn has the potential for further investments, which has the potential to drive the firm towards an IPO/ acquisition. Table 4-3 summarizes the changes made to the original dataset via filtering to analyze only those firms with investments.

Table 4-3 General vs Investment Representation

Type	General	Represented in Investment Data	Percent
Firms	3839	997	25.9%
Investors	1969	1956	99.3%
Investors - Individuals	418	418	100.0%
Investors - Companies	1551	1538	99.2%
Total	**5808**	**2953**	**50.8%**

Source: Author's Calculations

Table 4-3 confirms that only a small subset of the firms (26%) received investments from other firms. This could indicate a trend of specialized, targeted investment in only those firms with the best of products or innovations rather than a more widespread investment into the industry as whole. The other interpretation of this data is that these companies are considered “anchor companies” and the other companies in the dataset were created as a result of these companies being invested in. Further analysis of this possible phenomenon will follow in Section 4.2 The investor number being almost equal is not surprising, as the individuals were brought into the dataset from the investment dataset, and a small reduction in investors between the datasets could be attributed to a mislabeling of the categories, or a brand-new investor with no investments. Since investor numbers are almost equivalent, the remainder of this section will focus on the differences in firms.

Table 4-3 provides an initial understanding of the landscape of the invested firms from a geographical perspective, but a key aspect in understanding these trends are what kinds of firms were interesting to investors, which can be seen in Table 4-4.

Table 4-4 Firms by Employee Count

Employee Count	Invested Firms	All Firms	Percentage
1-10	312	917	34.0%
11-50	326	1,564	20.8%
51-100	43	254	16.9%
101-250	28	255	10.9%
251-500	8	119	6.7%
501-1000	1	26	3.8%
1001-5000	4	20	20.0%
Missing Data	275	684	40.2%
Total	**997**	**3839**	

Source: Author's Calculations

Missing data accounts for 40% of the data in Table 4-4, and this has the potential to weaken analysis. Based on the trends regarding employment, there is once again an expectation that the distribution of the missing data be primarily within the 1-50 range, and as such, the expectation is that this missing data will increase the overall scale of the results, rather than the results themselves.

Table 4-4 demonstrates investors are at two ends of the spectrum. On the one hand, there are higher proportions of small firms with investors. These firms are typically considered start-ups, and they are often the true disruptors and innovators in the field. As seen in the literature, investors invest in these companies if the idea or innovation is one that they believe will lead to large return on investments. Furthermore, investment in these smaller companies, while considered high risk in the sense of the probability of losing money, is actually fairly low risk, as the amount of money that needs to be raised in the funding rounds is comparatively small. Compounding this with a diversification of portfolio, the overall risk in investment into small firms is low. As the employee counts rise, the percent of invested firms goes down increasingly.

Additionally, investments into start-ups are typically followed with a strong presence of the investor. In this way, the investor typically allows access to their networks and may mentor the firm. This practice is beneficial for both the investor and the firm. By having a more hands-on approach, the investor is able to keep an eye on the investment in a sense and have a more direct impact on the outcome. From the firm's perspective, this additional access to resources, both monetary, technological, and perhaps even human is a great boon. By interfacing with the investor, the firm is able to gain knowledge faster and perhaps more efficiently than by themselves. The caveat here is that an investor must allow for some autonomy of the firm, otherwise multiple challenges will be had. As the number of employees increase, the firms are seen as more successful, and investors are less likely to invest in firms that are larger because they cannot directly influence them as easily and larger firms already have the capital and resources needed in order to be successful. On the opposite side of the employee counts, the largest companies and the most established received investments at a percentage rivaling that of 11-50, as the investment into this bracket could be considered as "too big to fail" and considered significantly safer for the investor.

4.1.3 Investment by Firm Specialization

Blockchain is a technology which can require relatively low upstart costs in terms of financial capital. It is fully feasible for a startup of 1-2 employees to pick up laptops and begin innovating the next blockchain application. The challenge is that situations like this typically utilize the technology for the most *obvious* or easiest application, as the costs are low and there are already resources available for a quicker startup. Investment into these firms does not need to be extremely large, as the upstart is relatively easy, and the tech is more simplistic, yet almost guarantees ROI. On the other hand, there are categories that need heavy investment, and these are typically those that utilize the technology in more advanced cases.

Blockchain is most often associated with cryptocurrencies. However, while cryptocurrencies are the best known, investors are investing in firms across a spectrum of application areas, as Table 4-5 shows.

Table 4-5 Investor and Firm category

Investor Category	#	%	Firm Category	#	%
finance	616	31.3%	commerce and shopping	234	23.5%
commerce and shopping	118	6.0%	software and analytics	77	7.7%
health	57	2.9%	finance	66	6.6%
sports and entertainment	51	2.6%	sports and entertainment	57	5.7%
software and analytics	35	1.8%	privacy and security	46	4.6%
hardware	28	1.4%	hardware	41	4.1%
property and real estate	19	1.0%	health	24	2.4%
community and lifestyle	16	0.8%	science and engineering	23	2.3%
natural resources and energy	16	0.8%	logistics and transportation	19	1.9%
personnel	16	0.8%	natural resources and energy	16	1.6%
science and engineering	16	0.8%	personnel	14	1.4%
education	13	0.7%	property and real estate	14	1.4%
logistics and transportation	12	0.6%	education	7	0.7%
privacy and security	12	0.6%	food and agriculture	6	0.6%
government and military	9	0.5%	mobile	5	0.5%
mobile	9	0.5%	community and lifestyle	4	0.4%
food and agriculture	6	0.3%	government and military	3	0.3%
missing data	920	46.7%	missing data	341	34.2%
Total	**1969**	**100.0%**		**997**	**100.0%**

Source: Author's Calculations

As can be seen in Table 4-5, financial applications of blockchain technologies account for only third in the most invested but are the category that most of the investment firms identify as. Table 4-5 faces similar challenges in terms of missing data as Table 4-4. Once again, based on application area analysis from the literature, the expectation is that the scale will change, but be unaltered at the top end. It would bring more detail to the lesser categories and strengthen the analysis. Examining the top 3 categories for each of the firm types reveals that blockchain is being used, as with a large number of technological innovations, for marketing and commerce purposes. Commerce and shopping, software and analytics and finance make up the large portion of the firm categories. From the investment side of things, it is not surprising that the number one is finance, as these are investment firms with large capital and mainly operate in the financial sector. Secondly, as stated above, a large subset of the technology is being utilized for marketing and commerce purposes, so it is to be expected that commerce firms will invest into the start-ups to ensure they are at the leading edge of the technology, allowing them to gain market share and entice their customers.

After finance and commerce applications, investors were most interested in Health [2.9%] and sport and entertainment [2.6%]. Both of these application areas are not primarily associated with blockchain yet have used blockchain technology to great effect. One major challenge in the medical community, and more specifically in the pharmaceutical community is the prevalence of drug counterfeiting , in which firms will sell a drug that may be "contaminated, or contain the wrong or no active ingredient" (US Food and Drug Administration, 2019). The practice is not only highly unethical but also extremely dangerous. It is for this reason that investors in the health space are turning to blockchain as a potential solution. Firms are utilizing the transaction-tracking power of blockchain to ensure that the drugs are authentic via deep learning, IoT and other advanced technologies, which aid in filtering out fake drugs.

Another category that is not often mentioned with respect to blockchain is sports and gaming. As seen in Table 4-5, firms specializing in *sports and entertainment* account for 5.7% of all firms. Similarly, in the age of influencer-driven online social networks (Chopra et al., 2021), multiple firms are utilizing the financial side of cryptocurrencies in order to encourage posting content in exchange for money. This is achieved via distributing view money (which is a form of crypto currency) to all viewers in the tree, encouraging the best content as the poster gets the most money. In essence, firms and investors in this space are mainly utilizing the "exclusivity" of a cryptocurrency-like system to drive marketing without being subject to the same restrictions and competition that normal marketing agencies face. These firms are the new-world loyalty agencies. By being loyal to the client's content, the end user is rewarded via the cryptocurrency.

Firms operating within the commerce and shopping category exhibit the highest percentage of investment. A majority of these firms focus on crypto-currency based loyalty programs to entice the consumer with financial rewards. Firms in the logistics spaces are utilizing the power of smart contracts to enable a more effective supply chain, leading to reduced costs and lower risk. From an investment perspective, the world is highly reliant on supply chains, and everyone is looking for the best way to reduce their risk and decrease prices to drive high consumer volumes. In turn, investors into this field are banking on the market wide adoption of the investee's technology, which would yield a very high ROI to the investing firm. This is similar in the transportation space, wherein investors are funding ideas and firms that have a high probability of mass adoption, leading to high ROIs. One such example of this is a firm that is utilizing smart contracts to establish efficient, fast, and safe traffic patterns. Should this technology gain a foothold, it will be utilized in multiple cities and locations, which will lead to high profitability of the firm, and thus bring large ROIs.

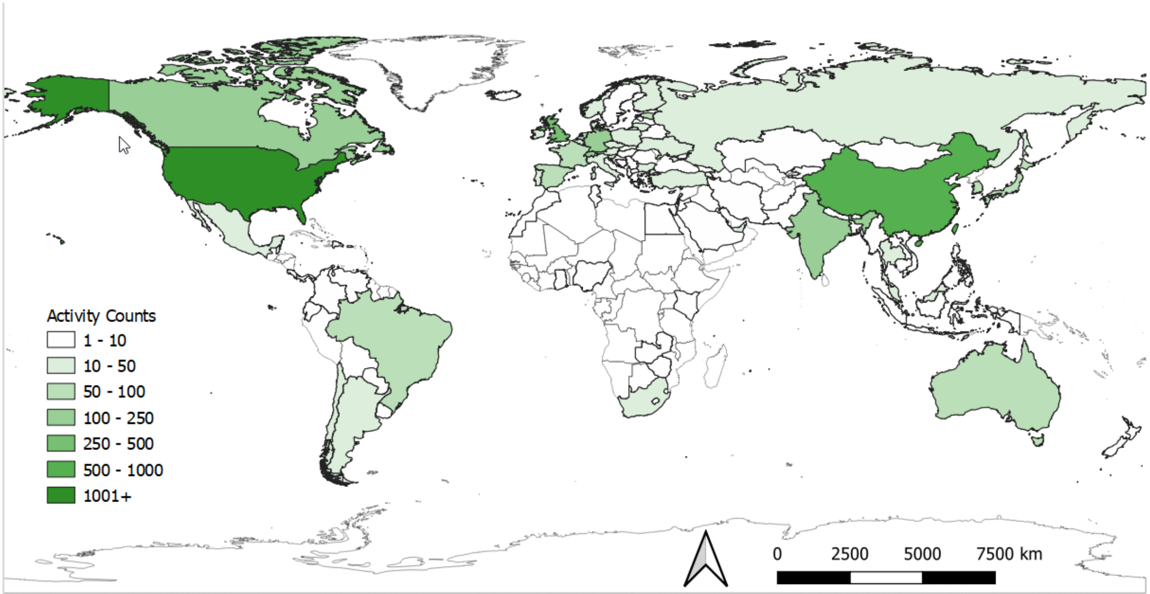
4.2 Geographic Patterns of the Industry

Scholars have long understood the role of geography in industry performance. For example, the notion of clustering, first introduced by Porter (1998) highlights the relationship between co-location and firm success. It is not surprising that certain places will have more of an influence on an industry; such places are typically associated with being economic, financial, and/or cultural capitals of the world. Within these locales, there are number of positive externalities from which firms can benefit, such as increased knowledge transfer, access to human and financial capital and access to global pipelines (Bathelt et al., 2004b). These locales provide strong start-up and innovation ecosystems, which in turn attract businesses. Relocation of firms, combined with the nurturing of homegrown talent through startups and spin-offs aid in firm attraction, which in turn causes a positive feedback loop. Supported by a strong ecosystem, these locales grow into potential clusters. Indeed, it is not only the firms, but the investors that will tend to locate in these places. Thus, in this section, the analysis aims to understand the global geography of the blockchain industry, both in terms of the location of blockchain firms and the patterns and networks of investment that support them. The measure for concentration in this section is unique firms by location. That is, the sum of uniquely named firms in the geographical area.

4.2.1 Global Geography of the Blockchain Industry

This section offers an understanding of the global geography of the industry. Based on observations made elsewhere in the literature, it is anticipated that blockchain firms would concentrate in world cities, as well as major technology hubs. Figure 4-3 shows a global map of the concentrations of blockchain firms and investors.

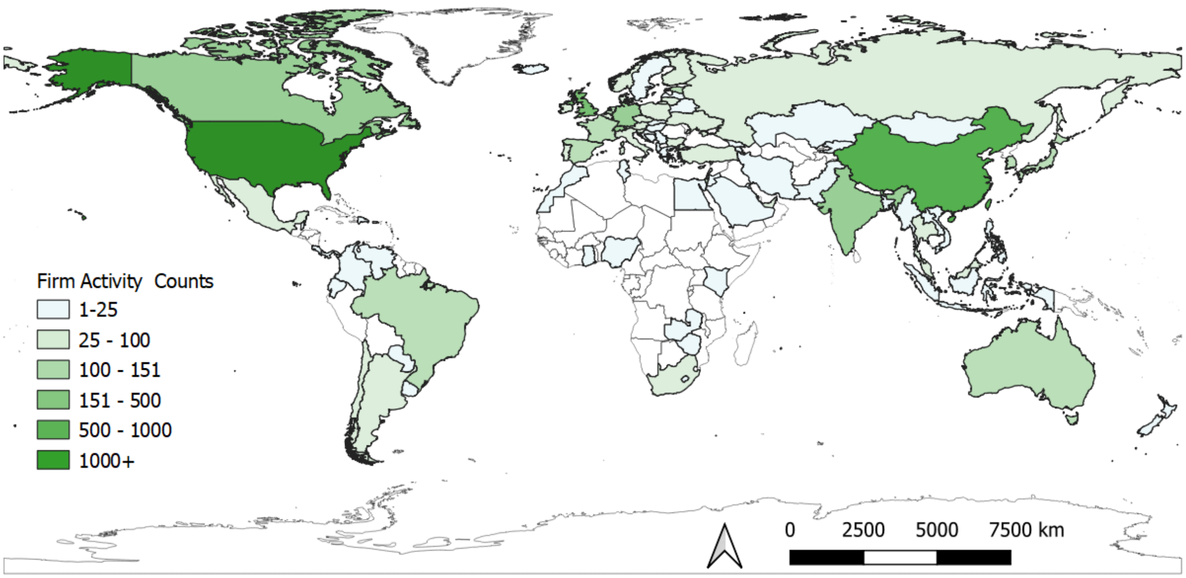
Figure 4-3 Country-level activity (Firms and Investors)



Source: Author's Calculations

Figure 4-3 demonstrates that blockchain activity (both firm and investor) can be found across the globe, but there are higher concentrations in some countries, namely the United States, China, and the United Kingdom. Each of these countries is home to a technology and/or financial capital: London (UK), Silicon Valley (US) and Beijing (China). To understand this pattern further, firms and investors are investigated separately.

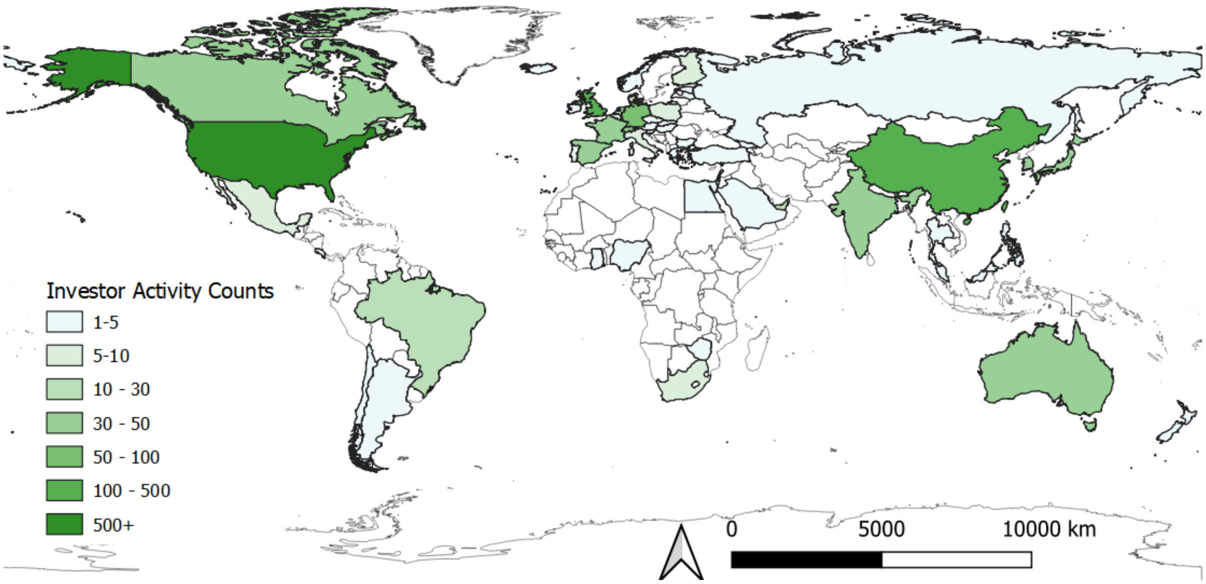
Figure 4-4 Country-level firm activity



Source: Author's Calculations

Figure 4-4 examines country-level firm activity across the globe, and it closely resembles the pattern seen in 4-3. The ecosystems and access to both financial and human capital within these locales both attracts firms to the countries and allows for the retention of local talent. Firms are globally distributed, yet exhibit strong co-location within certain countries, such as China, US, and England. These countries are able to provide incumbent firms with the needed resources to thrive, with access to the pipelines and the investors within close geographical proximity.

Figure 4-5 Country-level investor activity



Source: Author's Calculations

Figure 4-5 examines where blockchain investors are located. It shows that investors are more concentrated than blockchain firms, with high concentrations in only a few countries: United States, China, and the UK. Indeed, these countries house major global cities that offer high concentrations of financial and human capital. Notably, Figure 4-5 tells a story of significantly more investors in the United States than other countries. With home to 'superstar' cities such as New York and Silicon Valley, which have been centers of innovation and finance for multiple decades, it is not surprising to see that the US appears to be the key to this industry (Brail, 2019; Florida & Hathaway, 2018). Based on this national-level analysis, there is an expectation that, at the city-level, there will be clusters of activity within these countries. Table 4-6 (firms) and Table 4-7 (investors) provides the data behind Figures 4-3 to 4-5

Table 4-6 Country level firm counts

Position	Country	# Firms	% [n=3,839]
1	United States	1,266	33.0%
2	United Kingdom	345	9.0%
3	Singapore	210	5.5%
4	China	183	4.8%
5	Canada	156	4.1%
6	Switzerland	143	3.7%
7	India	137	3.6%
8	Germany	116	3.0%
9	Hong Kong	112	2.9%
10	Australia	74	1.9%
11	Netherlands	73	1.90%
12	Estonia	70	1.80%
13	France	60	1.60%
14	Spain	59	1.50%
15	Israel	47	1.20%

Source: Author's Calculations

Table 4-7 Country level investor counts

Position	Country	# Investors	% [n=1969]
1	United States	983	49.9%
2	China	197	10.0%
3	United Kingdom	134	6.8%
4	Germany	53	2.7%
5	Canada	49	2.5%
6	Hong Kong	47	2.4%
7	Japan	42	2.1%
8	India	35	1.8%
9	Singapore	35	1.8%
10	Switzerland	33	1.7%
11	France	32	1.60%
12	South Korea	32	1.60%
13	Australia	28	1.40%
14	Spain	26	1.30%
15	Israel	23	1.20%

Source: Author's Calculations

As expected, the Top 10 countries are countries that have been documented to be at the forefront of innovation and technology. The common theme in these countries is the high number of investors. When investors are present and investing into their own backyard, clusters tend to be created and prosper. This in turn creates a positive feedback loop in which firms receive funding from the investors, which in turn leads to more firms moving into the area, which further propels the area further. Another commonality within this list is the presence of world cities and technology capitals in a majority of the countries in Tables 4-6 and 4-7. These are countries with large capital and high GDP. Both firms and investors need access to capital and global pipelines, so this list is to be expected. Based on this table, the blockchain industry is highly globalized, with each part of the world having firms and investors. A notable difference between the US and others is visible in the numbers, but this could be due to data availability and not necessarily indicate the overwhelming position of the US as presented here. However, the domination of the United States in the technology space is not exactly new, as the country has been a mainstay in the technology and finance space for multiple generations, highlighted by strong institutional support, access to capital and high-technology areas such as Silicon Valley and New York. The largest difference between the US and most other countries, with the exception of China is the small difference between the number of investors and firms.

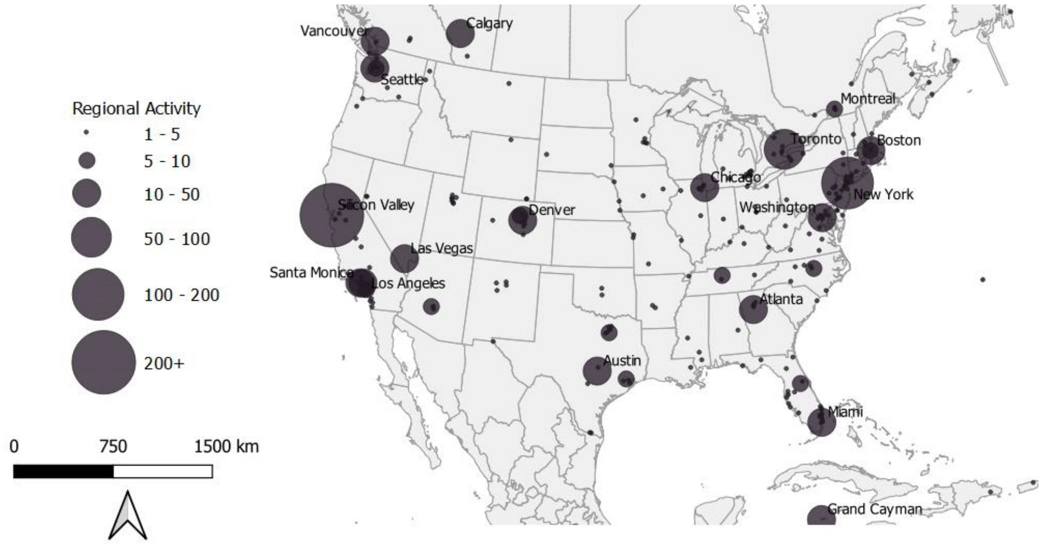
Sub-national (city-region) location patterns of blockchain firms in North America, Europe and Asia are examined. Additionally, analysis regarding possible regional specialization within application areas is presented for each region. Percentages in this section represent the percentage of total firms in the region as to the total number of firms in the application area. For example, if there are 100 unique firms in the application area, a city that hosts 10 unique firms in the application area would account for 10% of the total. Categories that follow the general patterns observed in the dataset of either low *n*, or primarily concentrated in the world cities were not subject to detailed analysis. Categories in this group include logistics and transportation, education, food, and agriculture. Associated figures may be found after the discussion.

Figure 4-6 shows the distribution of blockchain firms across North America. North America demonstrates clear activity clustering on both sides of the border. Silicon Valley and New York are the main hotspots of activity in this sub-region, but there is activity throughout. While overshadowed by the main hotspots, smaller clusters in Vancouver, Toronto and Seattle appear in the map. Categorical analysis of regions reveals that Montreal and Austin appear to be specialized in the software and analytics category, as they have little representation in the other categories. Specifically, in the case of Montreal, this is the only category for which this city is listed, with a 1.2% share of total unique firms. Sports and entertainment in North America is concentrated in known hotspots such as Los Angeles and the Cayman Islands. These are cities with a clear sports and entertainment cluster, with Hollywood in Los Angeles serving as a capital of the application area. Science and Engineering, which falls into the space of industrial application for this work, shows up in the major cities as per the other categories, but does show up in specific cities such as Austin, Chicago, and Johannesburg, which are cities without much representation in these categories, indicating potential regional specializations. Similarly, the application area of Science and Engineering appears to have concentrations within Austin [2.7%] and Chicago [2.7%]. Personnel, a category which is heavily represented by consulting firms, is extremely predominant in the world cities of the West, and more so with a strong national specialization in the United States. Within the top 15 cities represented in this category, 4 of them are located in the United States, and 6 of them are located in North America. Apart from Silicon Valley and New York, Atlanta [2.6%], Toronto [2.6%] and Seattle [1.7%] appear to have concentrations of personnel activity areas. The hardware application area is prevalent in the major cities but appears to be concentrated within San Diego [2.8%] and Raleigh [1.8%] in the United States. Nashville in particular holds a strong Health IT industry [3.9%], with several anchor institutions that promote a tech-forward approach to healthcare. In turn, this leads to advanced tech firms that wish to specialize in the field to co-locate in the area in order to benefit from the infrastructure and capital in the region (see Andes et al., 2016)). Property and real estate shows a clear regional specialization within Florida, as both Orlando and Miami are present in that list, with a total of 4% of the total global firms between them. Natural resources and Energy, an industry that hold a strong identity within certain cities and regions, yield regional specialization too, with Houston playing host to about 5% of the total [n=67] firms in the dataset (for a more in-depth analysis of Houston and the Energy industry, see Tutino et al., 2019). North America is diverse in application areas yet does display signs of regional specialization.

Figure 4-7 shows firm activity within Europe. There are high concentrations of firms in London, Amsterdam, as well Zug and Tallin. These cities are globally connected, with a number of investors and pre-existing tech ecosystems, either through the presence of technology investors (London), or the presence of strong anchor firms (Tallin and Zug). However, it appears that there are concentrations of blockchain firms present across European cities. It is possible that this is due to the stronger emphasis on cluster-based innovation policies and relevant institutional involvement and policy to encourage clustering in cities across different national contexts (Crescenzi et al., 2007). This is somewhat different from North America, where smaller clusters are very clearly overshadowed by Silicon Valley and New York. Firm activity within Europe is widespread over the categories. World cities within Europe such as London, Berlin, and Amsterdam can be found throughout the categorical analysis. Of note are cities such as Zug and Tallinn, as they appear in a multitude of categories, which speaks to the growth and maturity of the cities. Within commerce and shopping, Moscow appears to display signs of regional specialization in the area, but this may be a factor of the low firm count in Moscow, and the prevalence of commerce and shopping in the dataset. Within the Personnel activity area, some cities appear that do not have representation in other categories. Cities such as Dublin, Ireland [1.7%], Zurich, Switzerland [2.6%], and Sofia, Bulgaria [1.7%] appear to host a concentration of firms specializing in this application area. Zug and Zurich show activity in Property and Real Estate, and can be interpreted as a regional specialization in Switzerland, which may be the result of a specialized proptech/Contech cluster brought on by institutional involvement in the region (*Proptech Switzerland*, n.d.). In the Mobile category, even with a potential small n challenge [n= 23], the city of Kiev, Ukraine appears, and is indeed the only time that the city appears. Upon further examination, the local mobile provider, Kyivstar has started to utilize advanced technology such as big data and has massively improved the network infrastructure in the region (Kyivstar, 2020). Once again, this can be seen as a potential anchor firm, around which these firms will co-locate in order to be privy to the positive externalities brought on by the presence of such a firm in the region. Kyivstar has displayed interest in advanced technology, and the firm is extremely well funded, so it is understandable that a firm in the mobile space would co-locate in the region. Overall, Europe displays a wide range of activity, and this activity is more spread out across the region as opposed to North America.

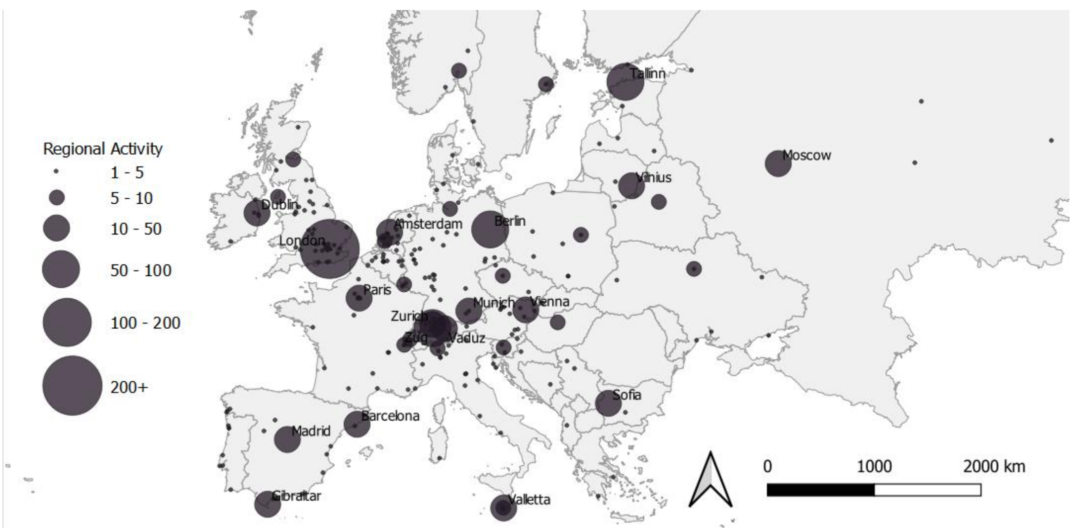
Figure 4-8 shows firm activity within Asia and Oceania. Tech clusters can be seen in the global cities in Asia: Shanghai, Beijing, Hong Kong, and Singapore. Smaller clusters in India, Israel, South Korea, and Japan are present. Regional specialization in this region can be seen in a small number of cities. Tel Aviv and Israel have been analyzed by economic thinktanks and they view Tel Aviv as a hotspot of tech innovation due to a strong economic infrastructure and concentration of technology within the city (Getz & Goldberg, 2016). Specifically, there is a concentration of software and analytics [1.8%] within Tel Aviv. Firms in the Hardware application area appear to have some concentration within India, with Chennai [1.8%] and Bangalore [1.8%] appearing, specifically in this category. Indeed, there is some scholarship to suggest that a hardware cluster is present in the region, with Chennai and Bangalore being top cities in the space (Khomiakova, 2007). MedTech appears to be prevalent across Asia, with both Taipei [3%] and Sydney [3%] appearing in the top 15 in the category. MedTech is a large industry in Australia, with a main cluster in Sydney.(NSW Treasury, 2019). Institutional involvement in the form of policy-driven innovation from the Taiwanese government, with an aim to create e-health could be a factor in the presence of Taipei in this category(European Union, 2018). Asia seems to exhibit similar trends to Europe, in that most major urban centers have some blockchain firms, especially in financial centers.

Figure 4-6 Regional activity - North America



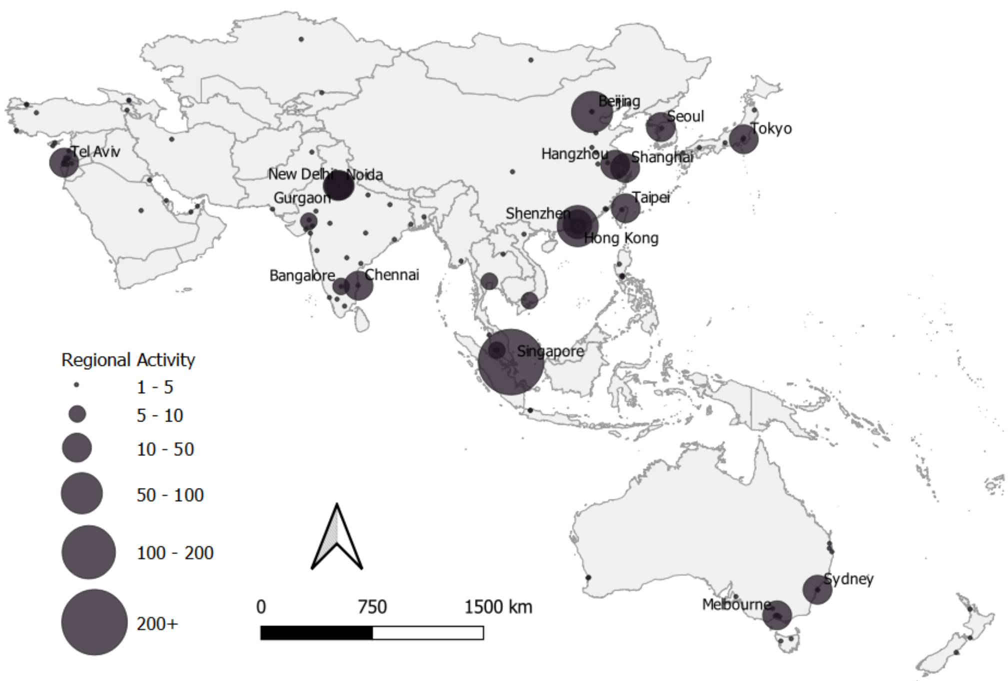
Source: Author's Calculations

Figure 4-7 Regional Activity - Europe



Source: Author's Calculations

Figure 4-8 Regional Activity - Asia and Oceania



Source: Author's Calculations

4.2.2 Geography of Investors

In addition to examining the location of blockchain firms, the world cities and clusters literature give a good prediction on the location of the investors. Investors need access to capital and networks. In addition, the literature suggests that investors will co-locate with firms, due to the ease of control and oversight. The locations of investor are shows in Table 4-8.

Table 4-8 Investor activity by region

Position	City Region	Country	# Investors	% [n=1969]
1	Silicon Valley	United States	441	22.4%
2	New York	United States	152	7.7%
3	London	United Kingdom	119	6.0%
4	Beijing	China	96	4.9%
5	Shanghai	China	43	2.2%
6	Tokyo	Japan	38	1.9%
7	Boston	United States	37	1.9%
8	Singapore	Singapore	35	1.8%
9	Los Angeles	United States	33	1.7%
10	Seoul	South Korea	27	1.4%
11	Toronto	Canada	27	1.4%
12	Berlin	Germany	25	1.3%
13	Chicago	United States	23	1.2%
14	Paris	France	23	1.2%
15	Hong Kong	Hong Kong	21	1.1%

Source: Author's Calculations

Blockchain investors are highly concentrated in capital and world cities. Established key players include the United States, United Kingdom and China, but trends show some other countries, such as Singapore and Japan, rising as strong investor capitals (Genberg, 2016). One item of interest in Table 4-8 is the large difference between investor activity in cities in the United States, especially Silicon Valley, and all other cities within the data set. This difference may be attributed to data availability and the aggregation of data for the super region of Silicon Valley. However, even without the creation of the super region, 3 out of the top 10 cities by investor count appear in the same geographical area [San Francisco, Palo Alto, Menlo Park]. These are areas of strong technological innovation that are anchored by tech giants such as Facebook and Google.

4.2.3 Geography of Firms

One key aspect of analysis is determining if blockchain investments are within-country, thus boosting the economy or if investments stretch beyond national borders, and if so, to which country(ies). The driving force of the economy has been found to be the cities. World cities are expected to be the main locations in the industry. This is due in part to the capital available to them, clustering, and access to human capital. A regional breakdown can be found in Table 4-9.

Table 4-9 Top city region firm activity

Position	City Region	Country	# Firms	% [n=3,839]
1	Silicon Valley	United States	347	9.0%
2	London	United Kingdom	286	7.5%
3	Singapore	Singapore	210	5.5%
4	New York	United States	184	4.8%
5	Beijing	China	82	2.1%
6	Zug	Switzerland	63	1.6%
7	Hong Kong	Hong Kong	60	1.6%
8	Toronto	Canada	60	1.6%
9	Tallinn	Estonia	58	1.5%
10	Berlin	Germany	56	1.5%
11	Los Angeles	United States	46	1.2%
12	Vancouver	Canada	41	1.1%
13	Amsterdam	Netherlands	36	0.9%
14	Paris	France	35	0.9%
15	Sydney	Australia	33	0.9%

Source: Author's Calculations

Overall, as demonstrated in Table 4-9, blockchain firms are less concentrated than the investors. Driving factors for the founding of firms in many of these countries can be linked with ease of access to capital, strong institutional support for new firms and access to human capital (Stam & van de Ven, 2019). For the most part, the city-regions on this list are viewed as tech superstar cities (Zukin, 2020), with the exception of Tallinn, Estonia and Zug, Switzerland. There are specific reasons that these two city-regions appear as important places for blockchain development. First, in Tallinn there is a strong anchor firm, Bolt, a platform ride hailing firm which was founded in the city due to tax incentives brought on by the economic policy of the region (Brail, 2019).Tallinn is also home to multinational tech firms such as Microsoft, Twilio and Wise, suggesting a very dynamic technology ecosystem. Zug, Switzerland has been dubbed “Crypto

Valley" (Crypto Valley Association, n.d.; Williams-Grut, 2018) as it is home to a number of blockchain companies, with one of the largest blockchain firms, the Ethereum Foundation, being located within the region. Ethereum is a blockchain-based cryptocurrency with a strong following and is the second most recognized cryptocurrency apart from Bitcoin. EF seeks to promote the use of Ethereum and other blockchain-related activities. In addition, Zug has established itself as a center of blockchain activity. The city was among the first in the world to allow payment in cryptocurrencies for traffic violations. The city is also host to large conferences surrounding blockchain. Lastly, the city is home to a number of blockchain unicorns. Along with this anchor firm, Zug emphasizes a strong economic policy, lax taxation on innovation, and access to world-class human capital (*Home*, n.d.). Thus, these two regions, while they are somewhat surprising places in discussions of the geography of innovation, indicate that institutional and policy support, both at the regional and national levels can encourage the emergence and growth of new technology clusters.

4.3 Geographic Patterns of Investment

In addition to examining the geography of the firms and investors, one of the chief goals of this thesis is to examine the relationships between the firms and the investors. In doing so, the relationships between places will be explored by examining the dynamics of firm-investor interactions. The analysis begins with a national level of investment dynamics, followed by a closer examination of city-level networks. Economic geography literature has long understood the need for domestic, or so-called "backyard" investment. This is investment into the region from within-region actors leading to a potentially virtuous cycle of investment and growth in addition, regions that grow are able to become potential clusters. The presence of this phenomenon will be examined within the dataset. One key aspect of a firm's ability to succeed is its ability to receive investment. Previous chapters have discussed the importance of being a world city or being located in a country with high global recognition, and this is where the analysis must begin.

4.3.1 Geographic Spread of Invested Firms

The previous analysis described the presence of world cities and their importance in the industry. These are places where a high proportion of firms and investors locate in order to take advantage of positive externalities. It follows that investors will either invest within the city, or invest elsewhere, within the top areas by activity. One key interest point is in what countries do firms get invested in the most. The results of this analysis will point to the cities and countries in which the majority of firms will want to settle in, or alternatively, other factors such as institutional involvement can be examined. From a firm perspective however, firms will want to settle into countries with the highest percentage of investment. Table 4-10 shows the differences between main and investment datasets based on the original Top 15.

Table 4-10 Levels of firm investment (%) by country

Position	Country	Total Firm Count	Invested Firm Count	% Invested
1	China	183	98	53.6%
2	South Korea	35	13	37.1%
3	Israel	47	15	31.9%
4	Singapore	210	57	27.1%
5	Japan	37	10	27.0%
6	United States	1266	324	25.6%
7	Australia	74	15	20.3%
8	United Kingdom	345	68	19.7%
9	Germany	116	22	19.0%
10	Switzerland	143	27	18.9%
11	Spain	59	11	18.6%
12	Hong Kong	112	20	17.9%
13	France	60	10	16.7%
14	Canada	156	20	12.8%
15	India	137	11	8.0%

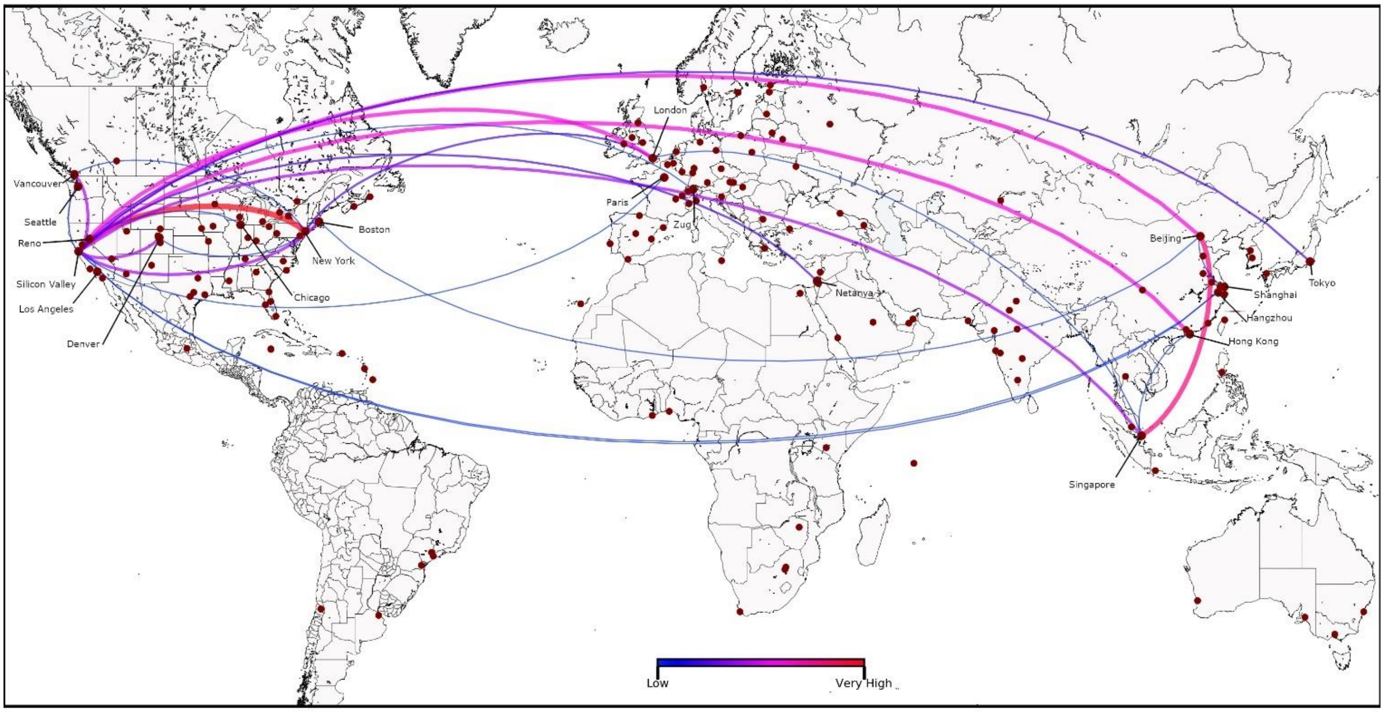
Source: Author's Calculations

Table 4-10 demonstrates a common trend in the data. In most countries, there is a large number of firms, but only a small subset receive investment. Indeed, the average for all countries in the top 10 apart from China is just over 23%. In most countries, only 1 in every 4 firms are actually funded on average. There are a few interesting cases from Table 4-10, being the US, Singapore, and China. Both US and Singapore boast high numbers of invested companies. As discussed in previous sections, these countries are centers of capital and innovation, so it is not unusual for the firms from these countries to receive more attention, brought on by pre-existing investor connections and the proliferation of a cluster in the areas. China, however, does show as a major outlier, with close to 53% of all created firms being invested. Apart from being a hotbed of innovation, financial and human capital, this result posits that firms in China are made with the investment already in mind. This would be a result of deep institutional support in the form of policy dedicated to getting investment, either national or international for qualifying firms. Such institutional support, along with increased access to human and financial capital would be reasons for the results as described in Table 4-10.

4.3.2 Network Analysis

In order to understand the deeper connections between cities, it is essential to quantify and visualize the interactions on a city-to-city basis. Social scientists have long use the avenue of network analysis to visualize and quantify relationships among objects (Golbeck, 2015), and in this case cities. Network analysis typically consists of data analysis, and then quantification of results. In this work, only exploratory network analysis will be performed. The analysis network is an undirected graph wherein the cities are represented by nodes, and the number of interactions between two cities is held in the weight between the nodes. As an example, if city A has 3 interactions to city B (A=>B: 3), and B=>A: 2, then the edge weight of AB is 5. This analysis will focus primarily on the eigenvector centrality of the nodes, in order to establish the key players in the industry. First, the data will be analyzed via looking at the node structure, followed by a community evaluation. Having established the main community, analysis will take part on this subset of data. Utilizing both intuition about contemporary economic geography literature and the results already presented in previous sections in this research, the community should be centered around the world cities and regions such as Silicon Valley and London. Additionally, this work has explored the potential of second tier world cities, or super-region capitals that are not the typical world cities, but through which a specific region’s interactions flow. In order to begin analysis, we must first get a general understanding of the network as a whole. The network consists of 463 nodes, with a total of 3687 edges. The main connected component is shown in Figure 4-9, with geographical accuracy.

Figure 4-9 The global geography of investment into blockchain firms[9]



Source: Author's Calculation

[9] This figure is currently under consideration for the Featured Graphic section at Environment and Planning A (see Appendix C)

As can be seen in Figure 4-9, the majority of interactions in the industry revolve around world cities, with very little interaction occurring outside of these cities. Cities such as Zug, Switzerland and Netanya, Israel [located 30 minutes north of Tel Aviv, which has been discussed as a rising force in the industry in this research] continue to feature prominently in this work as examples of cities that could be considered up and coming centers of innovation, spurred on by strong institutional involvement. This main network has an average weighted degree of 72.4, which means these are the main interactions of this network. When analyzing a network graph in order to understand the relationships, centrality is a well-accepted metric. As discussed in Chapter 3, this analysis using eigenvector centrality, but there are additional centrality measures that can be applied and will aid in supporting the findings of the eigenvector centrality. Given the fact that this analysis aims to understand the interactions among cities, one excellent measure of a node's (city's) importance to the network is betweenness centrality, which examines how often the node appears on the shortest path between the other nodes in the network. In this case, the more betweenness centrality a city has, the more knowledge that will be flowing through it, and these cities will be those with access to the strongest global pipelines (Golbeck, 2015). The betweenness centrality of the filtered component nodes with betweenness centrality greater than 0 is given in Table 4-11.

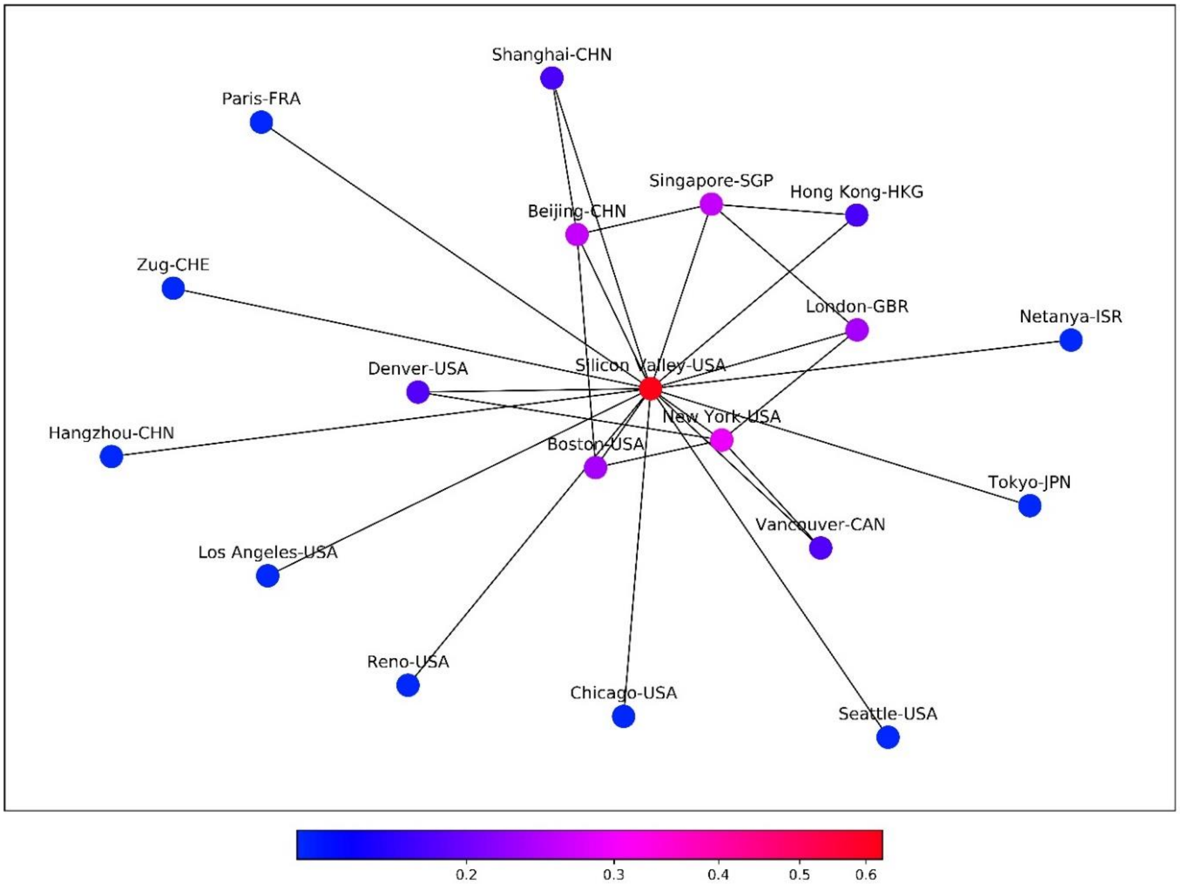
Table 4-11 Betweenness Centrality

City Region	Country	Between Centrality [0-1]
Silicon Valley	United States	0.895
New York	United States	0.020
Singapore	Singapore	0.010
Beijing	China	0.010
London	United Kingdom	0.003
Boston	United States	0.003

Source: Author's Calculations

Table 4-11 examines cities that were present on the shortest path between other cities. High betweenness centrality demonstrates that a large portion of the combined knowledge within the network passed through the city, as these cities were often a bridge between regions. Silicon Valley being at the top place with a betweenness centrality of 0.89 cements its importance to the network. The knowledge pipelines that are drivers of innovation will flow through Silicon Valley, and incumbent firms will be able to utilize the knowledge to increase competitiveness. Additionally, this may have a pull factor for firms, as the firms that wish to be most competitive in the space will aim to locate in those locales with the most knowledge and connections. The importance of Silicon Valley is further shown in Figure 4-10, which shows the eigenvector centrality.

Figure 4-10 Eigenvector centrality



Source: Author's Calculations

Figure 4-10 continues to provide evidence supporting contemporary notions within the economic geography literature surrounding the presence of global capitals and their importance to a global network. Cities such as Silicon Valley, Beijing and New York continue to be quantifiably of great importance to the network as a whole. There is a significant difference in the values of eigenvector centrality within these nodes, with Silicon Valley being quite far above all the other cities. This is not a surprising result, given the historical and current importance of Silicon Valley as a center of innovation and technology. New York and Beijing are both of extreme importance, secondary to Silicon Valley, and once again, these are cities with massive human and financial capital, combined with support infrastructure and access to global pipelines. The remainder of the cities in this list are well known hubs of innovation and capital, albeit at a lower scale as the main players. These are locales of great regional importance and serve as a gateway for their region to access the global networks.

The interactions on the city level in this industry are of key importance. This analysis continues to shed light on the importance of being a world city, but it also begins to explain the potential reasoning behind the success of some cities that were only seen as successful based on the presence of anchor firms and institutional involvement. Indeed, in addition to this, it is imperative for up-and-coming cities such as Zug, Switzerland and Tel Aviv, Israel to be connected to the world leaders in the industry. The analysis began with an introduction to the nodes and edges of this dataset and continued with an exploratory analysis of both betweenness centrality and eigenvector centrality. Betweenness centrality was highly biased towards Silicon Valley, which when combined with a similar result in the eigenvector centrality of 0.6, demonstrates the importance of this region to the global network. The major takeaway from this analysis is that regardless of institutional involvement, anchor firms and regional investment, or specializations, cities outside this network will struggle to maintain a large industry cluster.

4.4 Summary

This chapter aimed to answer the primary research question: What is the emerging geography of the blockchain industry? It did so by analyzing three different aspects of the dynamics of the blockchain industry. First, the nature of the blockchain industry was examined through the use of industry metrics and application areas. This section sought to answer the question: Does the blockchain industry conform to understood emerging tech patterns? The analysis reveals that the blockchain industry is comprised of relatively new firms with few employees; these firms are primarily active in high demand application areas such as commerce and shopping, software, and analytics. Most of the firms in the industry were founded post-2010 with high levels of growth up to 2017. Beginning in 2017, there was greater regulation and scrutiny of the industry, which appeared to contribute to the reduction in firm founding. This was chiefly due to the banning of ICOs, and the large amount of legal and financial regulations placed on the technology. As expected with new firms, employee counts were low, with a large portion being between 1-50 employees. Application areas followed expectations, with service sector applications being predominant. This did not detract from innovation brought in other industries such as logistics and health, for which specialized applications were described. Overall, the blockchain industry does seem to conform to the understood emerging tech patterns.

Geographical dynamics were the focus of the second section, in which the firms and investors were examined at both the national and city levels. This section sought to understand the spread and the key geographies of the blockchain industry. By incorporating the geography of the firms and investors, the role of countries, cities and regions could be examined. Overall, the analysis indicated that certain locales were of major interest to firms and investors, due to their human and financial capital. Additionally, the presence of positive externalities such as tacit knowledge transfer and access to global pipelines appear to be potential attraction factors. The analysis began with a high-level geographical overview, in which the key nations were found based on their total activity, in addition to specific firm and investor activity. In all cases, global leaders such as the United States, China and Singapore consistently ranked high in terms of activity. This may be due to pre-existing infrastructure within these regions, which allows firms to have access to investors, human and financial capital, and the global network. Next, the analysis narrowed to a city level. As expected, the top cities were located in the top countries, with Silicon Valley, London and New York consistently ranking at the top. In the literature, these are considered to be world cities and leading global tech hubs, so this work contributes favorably to the discussion on the importance of world cities in the emerging tech landscape (Brail, 2020).

Literature in economic geography examines the role of anchor firms and institutional involvement in supporting a subset of cities that have the potential to become world cities. This work has provided support for some of these claims in cities. For example, cities such as Zug, Switzerland and Tallinn, Estonia can be considered up-and-coming cities in the overall global landscape. Their ability to grow relies on the presence of an anchor firm, which is brought in by regional economic policy aimed at attracting firms into the region. With an anchor firm in the region, literature points to the development of the region into a potential cluster (Agrawal & Cockburn, 2003; Spencer, 2013; Zwick, 2017), bringing in much of the positive externalities that clusters provide. Additionally, analysis of the relationship between cities and category distributions was conducted in order to detect the presence of regional specialization. This was able to further cement the role of the world cities in the industry, with a large portion of each category being located in specific, well documented cities. In addition to these cities, potential regional specializations in application areas such as health [Nashville, US] and Natural Resources [Houston, US] were found. These results were then cross-examined with literature about the region, and in most cases, it was found that the present regional specialization attracted firms in the blockchain industry which wished to specialize in that application area.

Investor-firm relationships were then explored, with the goal of understanding the geography of investments and by doing so, examined the relationships among the cities and countries in the dataset. The investment profile of certain countries and cities aimed to explore those locales with the highest conversion percentage. At the country level, Asian countries, with a focus on China and South Korea showed the highest investment conversion.

In order to better quantify these relationships on a global scale, exploratory network analysis was conducted, with a focus on eigenvector centrality and betweenness, in order to verify earlier claims about the top locales being the most important to the network Quantitatively, Silicon Valley is the key to this network, with extremely high values in both betweenness and eigenvector centrality. For the most part, Silicon Valley has global influence and global information coursing within the region, falling in line with previous literature (Florida, 2019). Thus, it can be concluded that the blockchain industry investment is concentrated in certain key geographies, with a high prevalence in world cities. The patterns exhibited by the dataset do appear to follow the investment literature, in that the key investment cities are the same, and the interactions between these cities can be attributed to the desire to network and gain knowledge from the centers of innovation.

Chapter 5
Conclusion

The aim of this thesis was to examine the emerging geography of the blockchain industry. Understanding the locational patterns of firms operating in this nascent industry provides insight into the economic geography of emerging advanced technologies across the globe. The key question in this study was**: What is the emerging geography of the blockchain industry**? In order to delve deeper into industry dynamics, location, and investment patterns, three additional guiding questions structured the inquiry:

1. Does the blockchain industry conform to our a priori expectations of the characteristics of new technology industries?;
2. How is the industry spread out globally and what are the key geographies of the industry?; and
3. How is VC investment into the industry spread out across the globe, and do the patterns overlap with those found in the geography of investment and finance literature?

The blockchain industry was chosen for this study as it is an emerging industry that has received little academic attention, particularly in the context of economic geography. Blockchain technology is considered to be a disruptive technology with wide application and has garnered extensive interest amongst tech industry observers and investors since its inception in 2009 (Nakamoto, 2009). Blockchain is a decentralized transaction tracking system that boasts several benefits over existing technologies. The main advantages of the system are transparency, robustness, auditability and security (Christidis & Devetsikiotis, 2016). Initially, the technology was proposed for the anonymization of currency, but its transparent and distributed nature has seen applications across many industries, especially through smart contracts. In essence, blockchain allows multiple parties to verify transactions, which leads to enhanced trust and aids in optimization of the supply chain.

Overall, this study contributes to scholarship on emerging tech patterns by examining firm and industry dynamics of the blockchain industry. To date, there has been limited scholarly study of the blockchain industry, particularly from the perspective of economic geography; most scholarship has focused on technical aspects of blockchain or has included in-depth case studies about the usage of the technology in the industry. In contrast, this study has examined the location and investment patterns of blockchain firms on a global scale. The study uses quantitative descriptive analysis to assess the existing literature regarding firm dynamics and offers insights into the global dynamics of emerging tech industries.

The literature review began with a discussion of entrepreneurial ecosystems. It showed a picture of quick adopting entrepreneurs, with relatively low financial capital requirements, high presence of VC interaction and service-based applications. The resulting expectations included high amounts of firm founding once the tech became mainly accessible, low employee counts due to small start-ups with a need of human capital more so than financial capital, and the main application area of the technology being within the service sector. Following this section, the literature on the geography of innovation was examined to provide an understanding of the current geographical patterns associated with tech-intensive industries. This section centered around a discussion of clusters and the dynamics that form and maintain a cluster. Firms co-located in clusters are often subject to a number of advantages that make them more competitive compared to firms located elsewhere. Firms in clusters benefit due to the presence of capital in the region, access to global pipelines and the dissemination of tacit knowledge within the locale (Bathelt et al., 2004a; Breschi & Malerba, 2001; Mudambi et al., 2018; Porter, 2006). Geographically, technology-oriented clusters are typically found within superstar cities, or cities that are well known as tech and financial hubs, such as New York, London and Singapore (Brail, 2019; Florida, 2017; Florida & Hathaway, 2018). Additionally, this literature notes that while interest in emerging tech is global, outside of the well-known superstar cities, certain conditions must be met for a city to house a cluster of dynamic firms. More specifically, cluster formation is strongly tied to multiple factors, the most relevant of which being the presence of anchor firms and institutional involvement. For example, Brail (2019) provides the specific example of Tallinn, Estonia, which is home to the headquarters of Bolt, a ride-hailing firm that operates mainly in Europe and Africa. The scholarship on the geography of innovation presents a global, yet clustered industry, with a concentration of firms within a small number of city regions. Highly connected locales and world cities such as Silicon Valley, New York, London, and Beijing appear to account for most of the industry activity. In addition to these locales, a smaller subset of secondary and tertiary cities also appear to be important locations. Activity within these cities is augmented by the presence of an anchor firm in the region, often brought about through regional innovation policy.

Finally, the literature review examined the role of venture capital and geographic patterns of investment. In order to scale and support their innovation-related activities, firms require capital alongside other resources, often facilitated through access to networks. An investor is typically well connected and able to aid firms by providing networking opportunities. Literature in this section focused on the presence of investors in the world cities, where they are highly connected to other investors and there is a high concentration of human and financial capital. The literature also suggests that there is often co-location among investors and their portfolio of investments in order to optimize these interactions; this is especially true at the start-up phase. However, the literature notes that foreign investment is also common and does promote domestic investment. The literature suggests that VC will be located within high capital and networked areas such as world cities and will participate in both highly localized and global investment. Overall, Chapter Two highlighted that tech-based industries are typically located in clusters in world cities, with the investors either being local or located in other global centers.

To explore the specific geography of the blockchain industry and its patterns of investment, a quantitative methodology was used. The methodology for this research included the sourcing, cleaning and creation of a novel dataset including firms, investors, and the investments. In addition to aggregation and mapping-based analysis techniques, exploratory network analysis was performed, which allowed the industry dynamics to be explored. The dataset contained both quantitative [i.e., location, firm size] and qualitative [i.e., firm description] information on the firms and the investors. Data was sourced from Crunchbase, a large public database of firm and investor profiles. Data from blockchain firms and their investors were identified from the large dataset of over 800,000 firms. This database of 5838 firms and investors was then uploaded to Elasticsearch, a distributed search and analytical engine. Elasticsearch was chosen due to the flexibility, customizability, and access to services such as dashboard creation and mapping functions. Custom algorithms were created and used throughout the analysis, including a categorization algorithm, which aimed to categorize firm activity based on the self-identified categories within author-created a tier system. Lastly, exploratory network analysis was introduced, which aimed to quantify the relationships between cities with a focus on connectivity and networking which the academic literature finds as integral to the industry and innovation as a whole. Eigenvector centrality was used to identify investment flows related to blockchain activity and their relative importance.

The results were structured to address the three guiding questions. The analysis showed that most firms in the blockchain industry have been founded recently, with 54.6% of all firms being founded between 2016 and 2017. The rise in firm foundings could be associated with the relative ease of finding funding throughout this time in the form of ICOs, which was a way to crowd-source funding. Regulatory shifts and the associated global spillovers from the 2018 ICO ban in China may account for a sharp decline in firm founding, speaking to the impact that regulators and policymakers may have on an industry. Firm size was found to be in line with expectations from the literature, with 91% of all firms having 1-50 employees. Combined with an analysis of application areas of the firms, these firms appear to be concentrated within low startup cost, high demand areas such as commerce and shopping, finance and software and analytics. The low cost of entry due in part to pre-existing frameworks and applications combined with a steady demand from firms encouraged low employee count startups to blossom. Based on these findings, the blockchain industry does appear to conform to a priori expectations of tech-based clusters and activity.

The next research question relates to the global spread of the blockchain industry, with a particular interest in its key geographies. The literature points to a global yet concentrated industry, with a primary focus on co-location and clustering within world cities. Co-location within world cities allows firms to innovate with a supporting infrastructure and ecosystem while being subject to the positive externalities which have been found to improve firm performance, including presence of capital, tacit knowledge transfer and access to global pipelines. Activity is primarily concentrated in the superstar cities, but smaller pockets of activity appear to be present in most nations. Cities with the highest activity continue to be Silicon Valley, London, and Singapore. In addition to blockchain concentrations emerging in leading global cities, there are also concentrations of blockchain firms emerging in second-tier cities. These concentrations appear to follow one of two patterns. First, there are sets of cities where blockchain firms have emerged due to the presence of strong local conditions, including the presence of anchor firms and key institutions. Such places, including Tel Aviv, Israel and Zug, Switzerland, have been highlighted favorably in the popular and business press as up-and-coming places for tech-oriented development. Second, there is a subset of cities where the nature of the emerging blockchain industry reflects the specialization and composition of the surrounding regional economy. This includes places like Houston and Nashville, United States known for oil and health tech, respectively. The results of this section speak to the importance of the world cities literature in relation to contemporary economic development patterns. Additionally, this section adds favorably to the discussion surrounding the importance of anchor firms and institutional involvement and provides further evidence to the discussion regarding the presence of secondary and tertiary cities. The results of this section demonstrate that the blockchain industry is spread throughout the globe, with concentrations in world cities and major technology hubs. Key locations appear to be Silicon Valley, London, Beijing, and Singapore at the city level, mirrored by the United States, United Kingdom and China and Singapore at the national level. A large majority of industry activity is concentrated in these geographies.

Building on the importance of the connectedness within the industry for both firms and investors, the last section aimed to examine the geography of investment into the industry. The literature identifies the importance of proximity and co-location of investor and portfolio in order to maximize performance. Additionally, the literature noted that the geography of investment is similar to that of tech, but with more concentration of investors in the major world cities, due to the presence of capital and the ability to network. Investor firms were found to be highly concentrated within world cities, and their investments were both hyper local and global, with the majority of investments being found in other hubs. Silicon Valley and the Bay Area is the most highly connected, dominant place, with connections to both global financial centers (e.g., New York, Tokyo, London, and Singapore), as well as smaller urban tech hubs in Canada, Israel and elsewhere. Five locations account for approximately half of the 3,243 investments (51.4%): Silicon Valley, New York, Singapore, London, and Beijing; these five city-regions are also highly connected to one another. In addition, there appear to be supra-regional networks in the United States and Asia, suggesting that while investment is global, technology firms and their investors continue to occupy regional worlds. The results of this section speak to the concentrated geography of investment, with the majority of the investor firms being co-located within world cities. The strength and importance of these cites, with a focus on hubs such as Silicon Valley, London and Singapore appear to overlap with the literature.

Overall, this work has shed light on an industry that has not been the subject of extensive scrutiny by economic geographers and is one whose analysis could inform our understanding of the emergence of new technology-based industries. Furthermore, this thesis highlights the utility of using alternative datasets, such as Crunchbase, and algorithmic programming and distributed analysis engines such as Elasticsearch to understanding firm dynamics. Policy implications from this work encourage policy makers to understand the importance of key geographies in this industry and the importance of connecting their portfolio(s) with firms and investors in these geographies. Additionally, policy makers should be cognizant of the impact of anchor firms for a tech cluster and examine the viability of avenues which may attract such a firm to their region should a tech cluster be an economic development goal.

This study is not without limitations, however. As noted in Chapter 3, the limitations can be grouped into four major categories: source bias, author bias, firm self-reporting and missing/incomplete data. First, the firm that hosts the data is located in the United States, and as such is potentially subject to over-representation of U.S based firms. Second, there have been a number of author inputs in this work, be it by the creation of algorithms and procedures for data cleaning to categorization of firm activity. Each of these areas are subject to author bias, in that the author made decisions that could affect the outcome the results, and these decisions may not be in line with another researcher's decisions. It is in the nature of self-reported data, such as the case of firm data, that the data may not accurately reflect real-world values. In this case, the self-reporting of firm application areas, which was then categorized based on these values, could be subject to this challenge. Firms could over or under report their application areas in an effort to appeal to investors or other interested parties. This in turn has the potential to skew the categorization of the firm activity and has a potential effect on the analysis. In an effort to combat this, previous scholarship regarding the accuracy of the data source was consulted. In addition to favorable accuracy results from scholarship, the usage of both automated and human verification by the data source was noted. Lastly, data was not complete in all cases and firms with missing data were excluded from the analysis. It is unknown whether such errors are fully random or introduce systematic bias into the analysis. Multiple substitution and imputation methods were considered, but the lack of locational data made this impossible. In line with missing and incomplete data, several avenues for analysis that could have provided further quantitative evidence to claims made by this work were not utilized. In particular, a lack of financial data meant that analysis with the financial flows would be subject to low-n and bias challenges and was thus disregarded. It is the hope that future work can address some of these limitations and extend scholarship.

This thesis offers a beginning point for further analysis. Main points of further examination stem from methodological choices and data challenges. The first expansion of this work could come in the form of additional network analysis. This analysis can utilize more sophisticated algorithms to understand the interactions with a directed approach, in order get a better understanding of the role of investors and their interactions with the portfolio. Additionally, evolutionary networks and more in-depth analysis such as true regional analysis could be undertaken to further the understanding of regional specialization. One key data limitation that was faced by this work was the lack of financial data, which could provide further analysis of the financial flows with specific dollar amounts. In this way, average funding amounts can be examined, and different types of application areas can be examined to see which gets the most funding, furthering our understanding of the industry dynamics. Further examinations, which were beyond the scope of this paper, could include analysis of firm founders, which are integral to the industry. Additionally, the dynamics of exits and acquisitions could provide a deeper understanding of the patterns at play. Lastly, specific case studies with an aim to understand the firms, institutions and specific dynamics could be conducted. This entire work can then be extended out to other emerging tech industries such as Artificial Intelligence and Virtual Reality.

Bibliography

Acs, Z., & Audretsch, D. (1987). Innovation, Market Structure, and Firm Size. *The Review of Economics and Statistics*, *69*(4), 567–574.

Agrawal, A., & Cockburn, I. (2003). The anchor tenant hypothesis: Exploring the role of large, local, R&D-intensive firms in regional innovation systems. *International Journal of Industrial Organization*, *21*(9), 1227–1253. https://doi.org/10.1016/S0167-7187(03)00081-X

Amin, A. (1989). Flexible Specialisation and Small Firms in Italy: Myths and Realities. *Antipode*, *21*(1), 13–34. https://doi.org/10.1111/j.1467-8330.1989.tb00177.x

Andes, S., Ng, J., Gatz, C., & Muro, M. (2016). *From health care capital to innovation hub: Positioning Nashville as a leader in health IT*. 44.

Angraal, S., Krumholz, H. M., & Schulz, W. L. (2017). Blockchain Technology: Applications in Heath Care. *Circulation: Cardiovascular Quality and Outcomes*, *10*(9), e003800. https://doi.org/10.1161/CIRCOUTCOMES.117.003800

Audretsch, D. B., Hafenstein, M., Kritikos, A. S., & Schiersch, A. (2018). Firm Size and Innovation in the Service Sector. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3299312

Backman, M., & Lööf, H. (2015). The geography of innovation and entrepreneurship. *The Annals of Regional Science; Heidelberg*, *55*(1), 1–6. http://dx.doi.org.proxy.lib.uwaterloo.ca/10.1007/s00168-015-0713-x

Barnes, A., Brake, C., & Perry, T. (n.d.). *Digital Voting with the use of Blockchain Technology*. 19.

Bartoletti, M., Bellomy, B., & Pompianu, L. (2019). A Journey into Bitcoin Metadata. *Journal of Grid Computing*, *17*(1), 3–22. https://doi.org/10.1007/s10723-019-09473-3

Bathelt, H., Malmberg, A., & Maskell, P. (2004a). Clusters and knowledge: Local buzz , global pipelines and the process of knowledge creation. *Human Geography*, *1*, 31–56.

Bathelt, H., Malmberg, A., & Maskell, P. (2004b). Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, *28*(1), 31–56. https://doi.org/10.1191/0309132504ph469oa

Baumann, J., & Kritikos, A. S. (2016). The Link between R&D, Innovation and Productivity: Are Micro Firms Different? *IZA Discussion Papers*, 48.

Beaulieu, E., Benarroch, M., & Gaisford, J. (2004). Trade barriers and wage inequality in a North–South model with technology-driven intra-industry trade. *Journal of Development Economics*, *75*(1), 113–136. https://doi.org/10.1016/j.jdeveco.2003.07.007

Beaverstock, J. V., Smith, R. G., & Taylor, P. J. (1999). A roster of world cities. *Cities*, *16*(6), 445–458. https://doi.org/10.1016/S0264-2751(99)00042-6

Beaverstock, J. V., Smith, R. G., & Taylor, P. J. (2017). *GaWC - The World According to GaWC 2016*. https://www.lboro.ac.uk/gawc/world2016t.html

Bellavitis, C., Cumming, D., & Vanacker, T. (2020). Ban, Boom, and Echo! Entrepreneurship and Initial Coin Offerings. *Entrepreneurship Theory and Practice*, 1042258720940114. https://doi.org/10.1177/1042258720940114

Benali, M., & Burlat, P. (2012). Modeling cooperation links within networks of firms: A case study. *Journal of Intelligent Manufacturing; London*, *23*(6), 2451–2457. http://dx.doi.org.proxy.lib.uwaterloo.ca/10.1007/s10845-011-0511-y

Bjørgum, Ø., & Sørheim, R. (2015). The funding of new technology firms in a pre-commercial industry – the role of smart capital. *Technology Analysis & Strategic Management*, *27*(3), 249–266. https://doi.org/10.1080/09537325.2014.971002

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. https://doi.org/10.1080/13691060903184803

Bollinger, L., Hope, K., & James M. Utterback. (1983). A review of literature and hypotheses on new technology-based firms. *Research Policy*, *12*(1), 1–14. https://doi.org/10.1016/0048-7333(83)90023-9

Boschma, R. (1997). New industries and windows of locational opportunity. A long-term analysis of Belgium. *Erdkunde*, *51*, 12–22. https://doi.org/10.3112/erdkunde.1997.01.02

Brail, S. (2019). *World Cities of Ride-Hailing*. University of Toronto. https://munkschool.utoronto.ca/ipl/files/2019/01/BrailS_world-cities-of-ride-hailing-working-paper_29JA2019.pdf

Brail, S., & Vinodrai, T. (2020). The elusive, inclusive city: Toronto at a crossroads. In S. Bunce, N. Livingstone, L. March, S. Moore, & A. Walks (Eds.), *Critical Dialogues of Urban Governance, Development and Activism* (pp. 38–53). UCL Press. http://www.jstor.org/stable/j.ctv13xps83.10

Brant Carson, Giulio Romanelli, Patricia Walsh, & Askhat Zhumaev. (2018, June 19). *Blockchain beyond the hype: What is the strategic business value?* https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/blockchain-beyond-the-hype-what-is-the-strategic-business-value#

Braunerhjelm, P., & Feldman, M. P. (2006). Cluster Genesis: Technology-Based Industrial Development. In *Cluster Genesis*. Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199207183.001.0001/acprof-9780199207183

Breschi, S., & Malerba, F. (2001). The Geography of Innovation and Economic Clustering: Some Introductory Notes. *Industrial and Corporate Change*, *10*(4), 817–833. https://doi.org/10.1093/icc/10.4.817

Bresnahan, T., Gambardella, A., & Gambardella, P. of E. and M. L. of E. and M. A. (2004). *Building High-Tech Clusters: Silicon Valley and Beyond*. Cambridge University Press.

Breznitz, D. (2014). Why Silicon Valley Shouldn't Be the Model for Innovation. *Harvard Business Review*, 0–4.

Breznitz, S. M. (2014). *The Fountain of Knowledge: The Role of Universities in Economic Development*. Stanford University Press. http://www.sup.org/books/title/?id=17811

Bülbül, Ş., & İnce, G. (2018). Blockchain-based Framework for Customer Loyalty Program. *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, 342–346. https://doi.org/10.1109/UBMK.2018.8566642

Cai, Y., & Zhu, D. (2016). Fraud detections for online businesses: A perspective from blockchain technology. *Financial Innovation*, *2*(1), 20. https://doi.org/10.1186/s40854-016-0039-4

Chemmanur, T. J., Krishnan, K., & Nandy, D. K. (2011). How Does Venture Capital Financing Improve Efficiency in Private Firms? A Look Beneath the Surface. *The Review of Financial Studies*, *24*(12), 4037–4090. https://doi.org/10.1093/rfs/hhr096

Chen, G., Xu, B., Lu, M., & Chen, N.-S. (2018). Exploring blockchain technology and its potential applications for education. *Smart Learning Environments*, *5*(1), 1–1. https://doi.org/10.1186/s40561-017-0050-x

Chen, H., Gompers, P., Kovner, A., & Lerner, J. (2010). Buy local? The geography of venture capital. *Journal of Urban Economics*, *67*(1), 90–102. https://doi.org/10.1016/j.jue.2009.09.013

Cheng, S., Zeng, B., & Huang, Y. Z. (2017). Research on application model of blockchain technology in distributed electricity market. *IOP Conference Series: Earth and Environmental Science*, *93*, 012065. https://doi.org/10.1088/1755-1315/93/1/012065

Chopra, A., Avhad, V., & Jaju, and S. (2021). Influencer Marketing: An Exploratory Study to Identify Antecedents of Consumer Behavior of Millennial. *Business Perspectives and Research*, *9*(1), 77–91. https://doi.org/10.1177/2278533720923486

Christidis, K., & Devetsikiotis, M. (2016). Blockchains and Smart Contracts for the Internet of Things. *IEEE Access*, *4*, 2292–2303. https://doi.org/10.1109/ACCESS.2016.2566339

Council, N. R. (2011). *Growing Innovation Clusters for American Prosperity: Summary of a Symposium*. https://doi.org/10.17226/12926

Crescenzi, R., Rodríguez-Pose, A., & Storper, M. (2007). The territorial dynamics of innovation: A Europe–United States comparative analysis. *Journal of Economic Geography*, *7*(6), 673–709. https://doi.org/10.1093/jeg/lbm030

Crypto Valley Association. (n.d.). *Home*. Crypto Valley Association. Retrieved June 17, 2021, from https://cryptovalley.swiss/

Czarnitzki, D., & Binz, H. L. (2008). *R&D Investment and Financing Constraints of Small and Medium-Sized Firms*. *08*, 35.

Delgado, M., Porter, M. E., & Stern, S. (2014). Clusters, convergence, and economic performance. *Research Policy*, *43*(10), 1785–1799. https://doi.org/10.1016/j.respol.2014.05.007

Delmar, F., & Shane, S. (2002). *What Firm Founders Do: A Longitudinal Study of the Start-Up Process* (SSRN Scholarly Paper ID 1497834). Social Science Research Network. https://papers.ssrn.com/abstract=1497834

Deloitte. (2020). Blockchain Trends 2020. *C-Suite Briefing*, 16.

Derudder, B. (2008). Mapping Global Urban Networks: A Decade of Empirical World Cities Research. *Geography Compass*, *2*(2), 559–574. https://doi.org/10.1111/j.1749-8198.2008.00102.x

Dessemontet, P., Kaufmann, V., & Jemelin, C. (2010). Switzerland as a Single Metropolitan Area? A Study of its Commuting Network. *Urban Studies*, *47*(13), 2785–2802.

DG Enterprise and Industry. (2007). *Innovation Clusters in Europe: A statistical analysis and overview of current policy support* (DG Enterprise and Industry Report). https://cluster.hse.ru/mirror/pubs/share/212169880

Dinh, T. T. A., Liu, R., Zhang, M., Chen, G., Ooi, B. C., & Wang, J. (2018). Untangling Blockchain: A Data Processing View of Blockchain Systems. *IEEE Transactions on Knowledge and Data Engineering*, *30*(7), 1366–1385. https://doi.org/10.1109/TKDE.2017.2781227

Drosatos, G., & Kaldoudi, E. (2019). Blockchain Applications in the Biomedical Domain: A Scoping Review. *Computational and Structural Biotechnology Journal*, *17*, 229–240. https://doi.org/10.1016/j.csbj.2019.01.010

Dutta, S., & Folta, T. B. (2016). A comparison of the effect of angels and venture capitalists on innovation and value creation. *Journal of Business Venturing*, *31*(1), 39–54. https://doi.org/10.1016/j.jbusvent.2015.08.003

Eelke M. Heemskerk, Frank W. Takes, Javier Garcia-Bernardo, & M. Jouke Huijzer. (2016). Where is the Global Corporate �lite? A Large-scale Network Study of Local and Nonlocal Interlocking Directorates. *Sociologica*, *2*, 0–0. https://doi.org/10.2383/85292

Elasticsearch. (2021). *Keyword type family | Elasticsearch Reference [7.10] | Elastic* [Learn/Docs/Elasticsearch/Reference/7.10]. https://www.elastic.co/guide/en/elasticsearch/reference/current/keyword.html

Etzkowitz, H. (2019). Is Silicon Valley a global model or unique anomaly? *Industry and Higher Education*, *33*(2), 83–95. https://doi.org/10.1177/0950422218817734

European Union. (2018). *D.3.2 -Preparatory Briefing on Taiwan* (p. 33). European Cluster Collaboration Platform. https://clustercollaboration.eu/sites/default/files/d3.2_prep_briefing_taiwan_20181030.pdf

Fagiolo, G., Reyes, J., & Schiavo, S. (2010). The evolution of the world trade web: A weighted-network analysis. *Journal of Evolutionary Economics*, *20*(4), 479–514. https://doi.org/10.1007/s00191-009-0160-x

Fernandez-Vazquez, S., Rosillo, R., De La Fuente, D., & Priore, P. (2019). Blockchain in FinTech: A Mapping Study. *Sustainability*, *11*(22), 6366. https://doi.org/10.3390/su11226366

Florida, R. (2017). *The new urban crisis: How our cities are increasing inequality, deepening segregation, and failing the middle class-- and what we can do about it /*. Basic Books.

Florida, R. (2019, August 8). The Benefits of High-Tech Job Growth Don't Trickle Down. *Bloomberg.Com*. https://www.bloomberg.com/news/articles/2019-08-08/low-wage-workers-lose-out-when-tech-jobs-gain

Florida, R., & Hathaway, I. (2018). Rise of the Global Startup City. *Centre for American Entrepreneurship*, 76.

Friedman, M. (1986). Economists and Economic Policy. *Economic Inquiry*, *24*(1), 1–10. https://doi.org/10.1111/j.1465-7295.1986.tb01793.x

Friedmann, J., & Wolff, G. (1982). World city formation: An agenda for research and action. *International Journal of Urban and Regional Research*, *6*(3), 309–344. https://doi.org/10.1111/j.1468-2427.1982.tb00384.x

Gatteschi, V., Lamberti, F., Demartini, C., Pranteda, C., & Santamaría, V. (2018). To Blockchain or Not to Blockchain: That Is the Question. *IT Professional*, *20*(2), 62–74. https://doi.org/10.1109/MITP.2018.021921652

Genberg, H. (2016). Capital market development and emergence of institutional investors in the Asia-Pacific region. *Asia-Pacific Development Journal*, *22*(2), 1–26. https://doi.org/10.18356/abfeb31b-en

Gertler, M. S., & Vinodrai, T. (2005). Anchors of creativity: How public universities create competitive and cohesive communities. In F. Iacobucci & C. J. Tuohy, *Taking public universities seriously* (pp. 293–314). University of Toronto Press.

Getz, D., & Goldberg, I. (2016). *Best Practices and Lessons Learned in ICT Sector Innovation: A Case Study of Israel*. World Bank, Washington, DC. https://doi.org/10.1596/23644

Glaeser, E. L., Resseger, M., & Tobio, K. (2009). Inequality in Cities*. *Journal of Regional Science*, *49*(4), 617–646. https://doi.org/10.1111/j.1467-9787.2009.00627.x

Golbeck, J. (2015). Chapter 21—Analyzing networks. In J. Golbeck (Ed.), *Introduction to Social Media Investigation* (pp. 221–235). Syngress. https://doi.org/10.1016/B978-0-12-801656-5.00021-4

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Gugler, P., & Keller, M. (2009). *The Economic Performance of Swiss Regions*. 39.

Guo, J. (2015). On the Spillovers between Patents and Innovation in Japan. *Journal of Economics and Finance*, *39*(3), 590.

Guo, R., & Zhou, N. (2016). Innovation capability and post-IPO performance. *Review of Quantitative Finance and Accounting; New York*, *46*(2), 335–357. http://dx.doi.org.proxy.lib.uwaterloo.ca/10.1007/s11156-014-0471-3

Hall, P. (1966). *The world cities*. McGraw-Hill.

Heiberg, J., Binz, C., & Truffer, B. (2020). The Geography of Technology Legitimation: How Multiscalar Institutional Dynamics Matter for Path Creation in Emerging Industries. *Economic Geography*, *96*(5), 470–498. https://doi.org/10.1080/00130095.2020.1842189

Hellmann, T. F., & Puri, M. (1999). The Interaction between Product Market and Financing Strategy: The Role of Venture Capital. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.173655

Henn, S., & Bathelt, H. (2018). Cross-local knowledge fertilization, cluster emergence, and the generation of buzz. *Industrial and Corporate Change*, *27*(3), 449–466. https://doi.org/10.1093/icc/dtx036

Hinata, H. (2011). The Role of Local Government in Software and ITES Offshoring in Dalian, China. In M. Ohara, M. Vijayabaskar, & H. Lin, *Industrial Dynamics in China and India: Firms, Clusters, and Different Growth Paths*. Palgrave Macmillan UK. https://doi.org/10.1057/9780230308305

Home. (n.d.). Crypto Valley Association. Retrieved April 21, 2020, from https://cryptovalley.swiss/

IBM. (2017). *Forward Together: Three ways blockchain explorers chart a new direction* (No. 19; Global C-Suite Study). IBM. https://www.ibm.com/downloads/cas/P3WAB790

Ilgen, P. T. L. (2013). *Hard Power, Soft Power and the Future of Transatlantic Relations*. Ashgate Publishing, Ltd.

Intarakumnerd, P., & Vang, J. (2006). Clusters and Innovation Systems in Asia. *Science, Technology and Society*, *11*(1), 1–7. https://doi.org/10.1177/097172180501100101

Jain, A., & Jain, C. (2019). Blockchain Hysteria: Adding “Blockchain” to Company’s Name. *Economics Letters*, *181*, 178–181.

Jalilifard, A., Caridá, V. F., Mansano, A. F., Cristo, R. S., & da Fonseca, F. P. C. (2021). Semantic Sensitive TF-IDF to Determine Word Relevance in Documents. *ArXiv:2001.09896 [Cs, Stat]*. https://doi.org/10.1007/978-981-33-6977-1

Johansson, B., & Lööf, H. (2008). INNOVATION ACTIVITIES EXPLAINED BY FIRM ATTRIBUTES AND LOCATION. *Economics of Innovation and New Technology*, *17*(6), 533–552.

Joshi, K. (2018). Managing the risks from high-tech Investments in India: Differential strategies of foreign and domestic venture capital firms. *Journal of Global Entrepreneurship Research*, *8*(1), 21. https://doi.org/10.1186/s40497-018-0106-6

Kaiser, S. (2017). *Revealed comparative advantage and network centrality* [KU Leuven]. https://bookdown.org/sekaiser/Thesis/network-centrality-and-revealed-comparative-advantage.html#centrality-of-countries-in-the-propagation-of-shocks-in-the-international-trade-network

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, *64*(5), 402–406. https://doi.org/10.4097/kjae.2013.64.5.402

Kang, J.-K., Li, Y., & Oh, S. (2020). *Venture Capital Coordination in Syndicates, Corporate Monitoring, and Firm Performance*. 71.

Katz, J., & Gartner, W. B. (1988). Properties of Emerging Organizations. *The Academy of Management Review*, *13*(3), 429. https://doi.org/10.2307/258090

Ketels, C., & Protsiv, S. (2021). Cluster presence and economic performance: A new look based on European data. *Regional Studies*, *55*(2), 208–220. https://doi.org/10.1080/00343404.2020.1792435

Khan, N., Li, S., Safdar, M., & Khan, Z. (2019). The Role of Entrepreneurial Strategy, Network Ties, Human and Financial Capital in New Venture Performance. *Journal of Risk and Financial Management*, *12*(1), 41. https://doi.org/10.3390/jrfm12010041

Khomiakova, T. (2007). Information Technology Clusters in India. *Transition Studies Review*, *14*(2), 355–378.

Kondova, D. G. (2018). *The "Crypto Nation" Switzerland 2018*. 10.

Kononenko, O., Baysal, O., Holmes, R., & Godfrey, M. W. (2014). Mining modern repositories with elasticsearch. *Proceedings of the 11th Working Conference on Mining Software Repositories*, 328–331. https://doi.org/10.1145/2597073.2597091

Kortum, S., & Lerner, J. (2000). Assessing the Contribution of Venture Capital to Innovation. *The RAND Journal of Economics*, *31*(4), 674. https://doi.org/10.2307/2696354

Kshetri, N., & Voas, J. (2018). Blockchain-Enabled E-Voting. *IEEE Software*, *35*(4), 95–99. https://doi.org/10.1109/MS.2018.2801546

Kyivstar. (2020, December 24). *On the way to technological transformation – the story of Kyivstar*. KyivPost. https://www.kyivpost.com/business-wire/on-the-way-to-technological-transformation-the-story-of-kyivstar.html

Lamb, G., Suddon, M., & Brookfield Institute for Innovation and Entrepreneurship. (2016). *The state of Canada's tech sector, 2016*. Brookfield Institute for Innovation + Entrepreneurship.

Lee, N. (2019). Inclusive Growth in cities: A sympathetic critique. *Regional Studies*, *53*(3), 424–434. https://doi.org/10.1080/00343404.2018.1476753

Lee, N., & Clarke, S. (2019). Do low-skilled workers gain from high-tech employment growth? High-technology multipliers, employment and wages in Britain. *Research Policy*, *48*(9), 103803. https://doi.org/10.1016/j.respol.2019.05.012

Leppälä, S. (2016). Theoretical perspectives on localized knowledge spillovers and agglomeration. *Papers in Regional Science*, *97*(3), 467–484. https://doi.org/10.1111/pirs.12256

Liao, J. (Jon), & Welsch, H. (2008). Patterns of venture gestation process: Exploring the differences between tech and non-tech nascent entrepreneurs. *The Journal of High Technology Management Research*, *19*(2), 103–113. https://doi.org/10.1016/j.hitech.2008.10.003

Lu, R., Reve, T., Huang, J., Jian, Z., & Chen, M. (2018). A Literature Review of Cluster Theory: Are Relations Among Clusters Important? *Journal of Economic Surveys*, *32*(4), 1201–1220. https://doi.org/10.1111/joes.12255

Mason, C., & Harrison, R. (2004). Does investing in technology-based firms involve higher risk? An exploratory study of the performance of technology and non-technology investments by business angels*. *Venture Capital*, *6*(4), 313–332. https://doi.org/10.1080/1369106042000286471

McKinsey Global Institute. (2018, October). *The superstar firms, sectors, and cities leading the global economy | McKinsey*. https://www.mckinsey.com/featured-insights/innovation-and-growth/superstars-the-dynamics-of-firms-sectors-and-cities-leading-the-global-economy

McNeill, D. (2016). Governing a city of unicorns: Technology capital and the urban politics of San Francisco. *Urban Geography*, *37*(4), 494–513. https://doi.org/10.1080/02723638.2016.1139868

Mingo, S., Morales, F., & Dau, L. A. (2018). The interplay of national distances and regional networks: Private equity investments in emerging markets. *Journal of International Business Studies; Basingstoke*, *49*(3), 371–386. http://dx.doi.org.proxy.lib.uwaterloo.ca/10.1057/s41267-017-0141-5

Moretti, E. (2004). Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions. *American Economic Review*, *94*(3), 656–690. https://doi.org/10.1257/0002828041464623

Mudambi, R., Narula, R., & Santangelo, G. D. (2018). Location, collocation and innovation by multinational enterprises: A research agenda. *Industry and Innovation*, *25*(3), 229–241. https://doi.org/10.1080/13662716.2017.1415135

Nakamoto, S. (2009). *Bitcoin: A Peer-to-Peer Electronic Cash System*. 9.

Naz, A., Niebuhr, A., & Peters, J. C. (2015). What's behind the disparities in firm innovation rates across regions? Evidence on composition and context effects. *The Annals of Regional Science; Heidelberg*, *55*(1), 131–156. http://dx.doi.org.proxy.lib.uwaterloo.ca/10.1007/s00168-015-0694-9

Niosi, J., & Zhegu, M. (2005). Aerospace Clusters: Local or Global Knowledge Spillovers? *Industry and Innovation*, *12*(1), 5–29.

NSW Treasury. (2019, July 5). *MedTech—Medical innovation | Invest in NSW* (Australia). A Centre for Medical Innovation; corporateName: [site name]. https://invest.nsw.gov.au/sector-opportunities/medtech

Ōhara, M., Vijayabaskar, M., & Lin, H. (Eds.). (2011). *Industrial dynamics in China and India: Firms, clusters, and different growth paths*. Palgrave Macmillan.

Pasquero, J. (2000). Regional Market Integration in North America and Corporate Social ManagementEmerging Governance Frameworks for Business and Public Policy. *Business & Society - BUS SOC*, *39*, 6–23. https://doi.org/10.1177/000765030003900103

Pierro, M. D. (2017). What Is the Blockchain? *Computing in Science Engineering*, *19*(5), 92–95. https://doi.org/10.1109/MCSE.2017.3421554

Piras, G., Postiglione, P., & Aroca, P. (2012). Specialization, R&D and Productivity Growth: Evidence from EU Regions. *Annals of Regional Science*, *49*(1), 35–51. http://dx.doi.org.proxy.lib.uwaterloo.ca/10.1007/s00168-010-0424-2

Porter, M. E. (1998). *Clusters and the New Economics of Competition*. 77–90.

Porter, M. E. (2006). The Economic Performance of Regions. In *Regional Competitiveness*. Routledge.

PricewaterhouseCoopers. (2018). *Blockchain is here. What's your next move?* PwC. https://www.pwc.com/gx/en/industries/technology/blockchain/blockchain-in-business.html

PropTech Switzerland. (n.d.). Technology and Innovation in Real Estate, Construction and Finance. Retrieved February 3, 2021, from https://www.proptechswitzerland.com/

Ramos, J. (2003). *Using TF-IDF to Determine Word Relevance in Document Queries*. /paper/Using-TF-IDF-to-Determine-Word-Relevance-in-Queries-Ramos/b3bf6373ff41a115197cb5b30e57830c16130c2c

Richard Florida. (2014). The Creative Class and Economic Development. *Economic Development Quarterly*, *28*(3), 196–205. https://doi.org/10.1177/0891242414541693

Sacco, P. L. (2017). The Essence of Smart Specialization: Local Economies as Computational Platforms [Review Essay]. *Regional Studies*, *51*(5), 814–816.

Sassen, S. (1991). *The Global City: New York, London, Tokyo*. Princeton University Press. http://www.jstor.org/stable/j.ctt2jc93q

Saxenian, AnnaLee. (1994). *Regional advantage: Culture and competition in Silicon Valley and Route 128*. Harvard University Press. http://www.hup.harvard.edu/catalog.php?isbn=9780674753402&content=reviews

Schückes, M., & Gutmann, T. (2020). Why do startups pursue initial coin offerings (ICOs)? The role of economic drivers and social identity on funding choice. *Small Business Economics*. https://doi.org/10.1007/s11187-020-00337-9

Senor, D., & Singer, S. (2011). *Start-Up Nation: The Story of Israel's Economic Miracle*. McClelland & Stewart.

Shah, N., Willick, D., & Mago, V. (2018). A framework for social media data analytics using Elasticsearch and Kibana. *Wireless Networks; New York*, 1–9. http://dx.doi.org.proxy.lib.uwaterloo.ca/10.1007/s11276-018-01896-2

Smith, R. G. (2014). Beyond the Global City Concept and the Myth of ‘Command and Control.’ *International Journal of Urban and Regional Research*, *38*(1), 98–115. https://doi.org/10.1111/1468-2427.12024

Spencer, G. M. (2013). *The economic impact of anchor firms and industrial clusters:an analysis of Canadian and American manufacturing firms and clusters*. University of Toronto. https://localideas.files.wordpress.com/2014/05/anchor-firms-and-clusters.pdf

Spencer, G. M., Vinodrai, T., Gertler, M. S., & Wolfe, D. A. (2010). Do Clusters Make a Difference? Defining and Assessing their Economic Performance. *Regional Studies*, *44*(6), 697–715. https://doi.org/10.1080/00343400903107736

Spigel, B. (2020). *Entrepreneurial Ecosystems: Theory, Practice, Futures*. Edward Elgar. https://doi.org/10.4337/9781788975933

Spigel, B., & Stam, E. (2018). Entrepreneurial ecosystems. *SAGE Handbook of Entrepreneurship and Small Business*, 21. https://doi.org/10.4135/9781473984080.n21

Spigel, B., & Vinodrai, T. (2020). Meeting its Waterloo? Recycling in entrepreneurial ecosystems after anchor firm collapse. *Entrepreneurship & Regional Development*, *0*(0), 1–22. https://doi.org/10.1080/08985626.2020.1734262

Stam, E. (2015). Entrepreneurial Ecosystems and Regional Policy: A Sympathetic Critique. *European Planning Studies*, *23*(9), 1759–1769. https://doi.org/10.1080/09654313.2015.1061484

Stam, E., & van de Ven, A. (2019). Entrepreneurial ecosystem elements. *Small Business Economics*, *56*(2), 809–832. https://doi.org/10.1007/s11187-019-00270-6

Sternberg, R., & Arndt, O. (2001). The Firm or the Region: What Determines the Innovation Behavior of European Firms? *Economic Geography*, *77*(4), 364–382. https://doi.org/10.1111/j.1944-8287.2001.tb00170.x

Sun, J., Yan, J., & Zhang, K. Z. K. (2016). Blockchain-based sharing services: What blockchain technology can contribute to smart cities. *Financial Innovation*, *2*(1), 26. https://doi.org/10.1186/s40854-016-0040-y

Sutton, S. M. (2000). The role of process in software start-up. *IEEE Software*, *17*(4), 33–39. https://doi.org/10.1109/52.854066

Tan, B., Yan, J., Chen, S., & Liu, X. (2018). The Impact of Blockchain on Food Supply Chain: The Case of Walmart. In M. Qiu (Ed.), *Smart Blockchain* (Vol. 11373, pp. 167–177). Springer International Publishing. https://doi.org/10.1007/978-3-030-05764-0_18

Tapscott, D., & Tapscott, A. (2017). How Blockchain Will Change Organizations. *MIT Sloan Management Review; Cambridge*, *58*(2), 10–13.

Tarasconi, G., & Menon, C. (2017). *Matching Crunchbase with patent data* (OECD Science, Technology and Industry Working Papers No. 2017/07; OECD Science, Technology and Industry Working Papers, Vol. 2017/07, p. 21). OECD. https://doi.org/10.1787/15f967fa-en

Teirlinck, P., & Khoshnevis, P. (2019). Within-cluster determinants of output efficiency of R&D in the space industry. *Omega*, 102039. https://doi.org/10.1016/j.omega.2019.02.008

The European Cluster Observatory: EU cluster mapping and strengthening clusters in Europe. (2009). Publications Office. https://data.europa.eu/doi/10.2769/10419

Tornikoski, E., & Renko, M. (2014). Timely Creation of New Organizations: The Imprinting Effects of Entrepreneurs' Initial Founding Decisions. *M@n@gement, Vol. 17*(3), 193–213.

Trujillo, J. L., & Parilla, J. (2016, September 29). Redefining Global Cities. *Brookings*. https://www.brookings.edu/research/redefining-global-cities/

Turkina, E., Van Assche, A., & Kali, R. (2016). Structure and evolution of global cluster networks: Evidence from the aerospace industry. *Journal of Economic Geography*, *16*(6), 1211–1234. https://doi.org/10.1093/jeg/lbw020

Tutino, J., Melosi, M. V., Joseph A. Pratt, & Martin V. Melosi. (2019). Energy Capital and Opportunity City: Houston in the Twentieth Century. In *New World Cities* (pp. 242–294). University of North Carolina Press. http://www.jstor.com/stable/10.5149/9781469648774_tutino.11

US Food and Drug Administration. (2019, September 13). *Counterfeit Medicine*. FDA; FDA. https://www.fda.gov/drugs/buying-using-medicine-safely/counterfeit-medicine

Wan, Y., & Liu, H. (2011). Research about relation of knowledge spillover and technology innovation. *2011 International Conference on Remote Sensing, Environment and Transportation Engineering*, 604–607. https://doi.org/10.1109/RSETE.2011.5964349

Wang, W., & Street, W. N. (2015). Modeling influence diffusion to uncover influence centrality and community structure in social networks. *Social Network Analysis and Mining*, *5*(1), 15. https://doi.org/10.1007/s13278-015-0254-4

Westhead, P., & Storey, D. (1997). Financial constraints on the growth of high technology small firms in the United Kingdom. *Applied Financial Economics*, *7*(2), 197–201.

Where is Silicon Valley? (n.d.). Silicon Valley Historical Association. Retrieved January 29, 2021, from https://www.siliconvalleyhistorical.org/where-is-silicon-valley

Williams-Grut, O. (2018, July 2). *Here's what it's like to visit 'Crypto Valley'—Switzerland's picturesque blockchain version of Silicon Valley*. Business Insider. https://www.businessinsider.com/what-its-like-in-zug-switzerlands-crypto-valley-2018-6

Wolf, M., & Terrell, D. (2016). *The high-tech industry, what is it and why it matters to our economic future*. 7.

Wolfe, D. A. & Conference Board of Canada. (2009). *21st century cities in Canada: The geography of innovation : the CIBC scholar-in-residence lecture*. Conference Board of Canada. https://central.bac-lac.gc.ca/.item?id=21st-Century-Cities-in-&op=pdf&app=Library

Wolfe, D., & Bramwell, A. (2016). Innovation, creativity and governance 8 years on: Social dynamics of economic performance in city-regions. *Innovation*, *18*(4), 462–467. https://doi.org/10.1080/14479338.2016.1243015

Wolf-Powers, L. (2005). Up-Zoning New York City's Mixed-Use Neighborhoods: Property-Led Economic Development and the Anatomy of a Planning Dilemma. *Journal of Planning Education and Research*, *24*(4), 379–393. https://doi.org/10.1177/0739456X04270125

Wu, D., Liu, X., Yan, X., Peng, R., & Li, G. (2019). Equilibrium analysis of bitcoin block withholding attack: A generalized model. *Reliability Engineering & System Safety*, *185*, 318–328. https://doi.org/10.1016/j.ress.2018.12.026

Wu, K., Ma, Y., Huang, G., & Liu, X. (2019). A First Look at Blockchain-based Decentralized Applications. *ArXiv:1909.00939 [Cs]*. http://arxiv.org/abs/1909.00939

Yang, Q., Xu, F., Zhang, Y., Liu, F., Hu, W., & Liao, Q. (2018). Design and Implementation of a Loan System Based on Smart Contract. In M. Qiu (Ed.), *Smart Blockchain* (Vol. 11373, pp. 22–31). Springer International Publishing. https://doi.org/10.1007/978-3-030-05764-0_3

Zhang, R., Xue, R., & Liu, L. (2019). Security and Privacy on Blockchain. *ACM Computing Surveys*, *52*(3), 51:1-51:34. https://doi.org/10.1145/3316481

Zook, M. A., & Blankenship, J. (2018). New spaces of disruption? The failures of Bitcoin and the rhetorical power of algorithmic governance. *Geoforum*, *96*, 248–255. https://doi.org/10.1016/j.geoforum.2018.08.023

Zukin, S. (2020). *The Innovation Complex: Cities, Tech, and the New Economy*. Oxford University Press.

Zwick, A. (2017). *Economic Specialization in Ontario's Mid-Sized Cities: Clusters and Anchor Institutions*. 7.

Appendix

Appendix A

Aggregation into Silicon Valley

Cities within Silicon Valley			
Alameda	Gilroy	Palo Alto	Saratoga
Atherton	Hayward	Pleasanton	Sausalito
Berkeley	Hillsborough	Portola Valley	Scotts Valley
Burlingame	Los Altos	Redwood City	South San Francisco
Concord	Los Gatos	Redwood Shores	Stanford
Corte Madera	Menlo Park	Richmond	Sunnyvale
Cupertino	Millbrae	San Bruno	Walnut Creek
Daly City	Milpitas	San Francisco	Woodside
Danville	Monte Sereno	San Jose	
Davis	Monterey	San Leandro	
Dublin	Moraga	San Mateo	
East Palo Alto	Mountain View	San Rafael	
Emeryville	Newark	San Ramon	
Foster City	Oakland	Santa Clara	
Fremont	Orinda	Santa Cruz	

Source : Silicon Valley Historical Association (*Where Is Silicon Valley?*, n.d.)

Appendix B

Categorization Mappings

Category	Assigned Category
Software	Software and Analytics
financial services	Finance
Payments	Commerce and shopping
information technology	Software and Analytics
internet services	Software and Analytics
data and analytics	Software and Analytics
Hardware	Hardware
privacy and security	Privacy and Security
lending and investments	Finance
artificial intelligence	Software and Analytics
commerce and shopping	Commerce and Shopping
science and engineering	Science and Engineering
mobile	Mobile
media and entertainment	Sports and Entertainment
apps	Software and Analytics
consumer electronics	Hardware
gaming	Sports and Entertainment
sales and marketing	Commerce and Shopping
professional services	Personnel
transportation	Logistics and Transportation
content and publishing	Sports and Entertainment
real estate	Property and Real Estate
health care	Health
community and lifestyle	Community and Lifestyle
design	Commerce and Shopping
advertising	Commerce and Shopping
natural resources	Natural Resources and Energy
messaging and telecommunications	Mobile
sports	Sports and Entertainment
government and military	Government and Military
video	Sports and Entertainment
education	Education
energy	Natural Resources and Energy
sustainability	Natural Resources and Energy
biotechnology	Health
music and audio	Sports and Entertainment
travel and tourism	Logistics and Transportation

agriculture and farming	Food and Agriculture
navigation and mapping	Logistics and Transportation
food and beverage	Food and Agriculture
platforms	Software and Analytics
administrative services	Personnel
manufacturing	Hardware
consumer goods	commerce and shopping
events	Sports and Entertainment
clothing and apparel	commerce and shopping

Appendix C

Featured Graphic Submitted to Environment and Planning A

The global geography of investment in emerging technologies: The case of blockchain firms

Martin Holicka
University of Waterloo, Canada
mholicka@uwaterloo.ca

Tara Vinodrai
University of Toronto, Canada
tara.vinodrai@utoronto.ca

Geographers have long been interested in where new technologies and industries emerge. Technology firms tend to locate in world cities and technology hubs, with concentrations of highly skilled workers, venture capital, anchor institutions and knowledge infrastructure (Brail, 2020). Blockchain is one such emerging technology, which emerged in 2009 as a distributed ledger technology that maintains transparency, security, and accountability in transactions. Best known as a disruptive financial technology associated with cryptocurrencies such as Bitcoin, blockchain has application in supply chain optimization, royalty and copyright tracking, cybersecurity, refugee identity and transaction systems, and voting systems. Recent studies of the blockchain industry have examined the role of cryptocurrencies in financialization, the relationship between venture capital and start-up dynamics, and firm competitiveness (Fernandez-Vazquez et al., 2019; Park et al., 2020; Zook and Grote, 2020). However, few studies focus explicitly on the geography of the blockchain industry or the investment flows supporting these firms.

We developed a database of 3,839 blockchain firms that were founded between 2010 and 2018 using Crunchbase, an accepted source for analyzing entrepreneurial firm dynamics (Block and Sandner, 2009; Friedlmaier et al., 2018). We included data on employment, location, and specializations. For firms receiving investments (25.9%), we collected information on investment rounds. We geocoded the dataset and created network data at the city-region level to capture investment interactions.

We visualize the global geography of blockchain firms and the related network of investment (Figure 1), showing places with ten or more investments to capture meaningful levels of interactions. We

illustrate the level of interaction between city-regions from low (blue) to very high (red), using a natural breaks scale determined by the Jenks optimization algorithm. Line widths accentuate differences in levels of interaction; thicker lines represent higher numbers of interactions.
While there are blockchain firms on almost every continent, firms receiving investments are concentrated in a much smaller number of city-regions. Not surprisingly, Silicon Valley and the Bay Area is the most highly connected, dominant place, with connections to both global financial centers (e.g., New York, Tokyo, London, and Singapore), as well as smaller urban tech hubs in Canada, Israel and elsewhere. Five locations account for approximately half of the 3,243 investments (51.4%): Silicon Valley, New York, Singapore, London, and Beijing; these five city-regions are also highly connected to one another. In addition, there appear to be supra-regional networks in the United States and Asia, suggesting that while investment is global, technology firms and their investors continue to occupy regional worlds. Overall, our analysis shows that while blockchain firms are located across the globe, the underlying patterns overlay with well-known patterns associated with the geographies of innovation.

References

Block J and 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. DOI: 10.1080/13691060903184803.

Brail S (2020) World cities of ride-hailing. *Urban Geography* DOI: 10.1080/02723638.2020.1775030

Fernandez-Vazquez S, Rosillo R, De La Fuente D, et al. (2019) Blockchain in FinTech: A mapping study. *Sustainability* 11(22): 6366. DOI: 10.3390/su11226366.

Friedlmaier M, Tumasjan A and Welpe IM (2018) Disrupting industries with blockchain: The industry, venture capital funding, and regional distribution of blockchain ventures. Available at: https://core.ac.uk/reader/143481280 (accessed 17 May 2021).

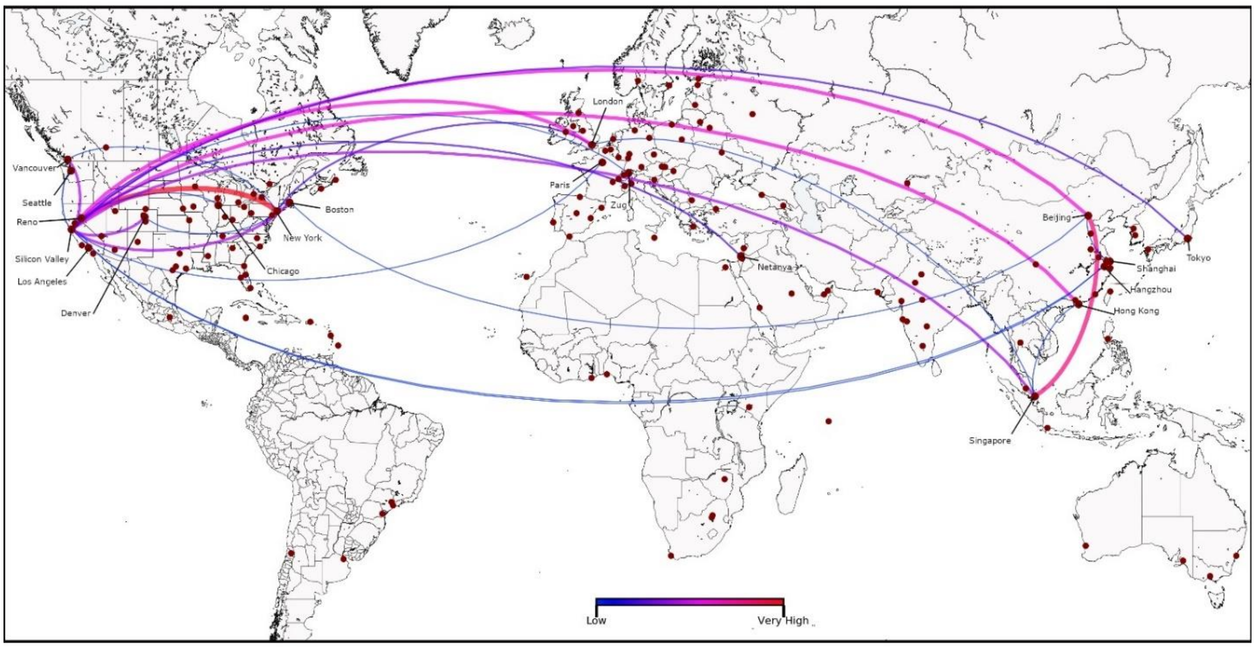
Park, G, Seungryul, RS and Choy, M (2020) Early mover (dis)advantages and knowledge spillover effects on blockchain startups' funding and innovation performance. *Journal of Business Research* 109: 64-75. https://doi.org/10.1016/j.jbusres.2019.11.068

Zook M and Grote MH (2020) Initial coin offerings: Linking technology and financialization. *Environment and Planning A: Economy and Space* 52(8): 1560–1582. DOI: 10.1177/0308518X20954440.

Acknowledgements

The authors would like to acknowledge funding and support from the Social Science and Humanities Research Council of Canada, as well as the Innovation Policy Lab (University of Toronto). The authors are grateful for the constructive comments offered by Shauna Brail (University of Toronto) and Heather Hall (University of Waterloo)

Figure 1: The global geography of investment in blockchain firms



Source: Crunchbase [Authors' calculations]