MECHANISM DESIGN ISSUES IN TECHNOLOGICAL SYSTEMS

by

Anparasan Mahalingam

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Industrial Engineering

West Lafayette, Indiana

August 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. David R. Johnson, Chair
School of Industrial Engineering

Dr. Mark R. Lehto
School of Industrial Engineering

Dr. Seungyoon Lee
Brian Lamb School of Communication

Dr. Jovan Grahovac
Krannert School of Management

Approved by:
Dr. Abhijit Deshmukh

To Appa and Amma – For shaping my intellectual journey.

ACKNOWLEDGMENTS

My PhDs (it still feels quite ironic to say PhD in plural) at Purdue University are a culmination of myriad learnings from so many people from all walks of life. For years, I had a vague idea that engineering and hard technological advancements will benefit from a systematic interface to both business strategy and public policy. Both my PhDs at Purdue are, in essence, operationalizing this idealistic and imprecise notion into specific coherent dissertations. Although, *ex post,* they look relatively cogent, my dissertation path had been messy, and as in the words of Bruce Springsteen it often felt like I, “[..] ain't learnin'; [..]; One step up and two steps back”.

This Industrial Engineering Ph.D. is truly a result of the rewarding camaraderie I have developed over the years with Prof. David Johnson. Since I first met David in 2015, his intellectual rigor in the area of mathematical simulation and modeling have shaped my research tremendously. I am indebted for his resolute mentorship, trust and friendship. I am also very grateful to the entire committee members, Profs. Mark Lehto, Seungyoon Lee and Jovan Grahovac, for providing valuable feedback and giving me their time generously to shape the thoughts for this dissertation, and broadly for my research agenda. There is an ancient mantra in South Asia on “guru” – teacher – that pronounces that a guru creates knowledge, preserves it and dispels the darkness of ignorance in a student. I am so humbled and honored to have had all these gurus during my dissertation process. I look forward to continuing to learn from all of them.

Growing up in Jaffna peninsula of Sri Lanka during the peak of civil war misfortunes were plenty, except for two: the abundance of intellectually sound education and a closely knit love-filled family. My parents inculcated the value of education in us; it was the single most important hope that guided all five of us brothers through the turbulent period. My father’s mental cookbook was that education combined with strong ethics will serve as the lighthouse for his sons in any stormy waters of life. While I will miss him dearly in all the years to come, I aspire to stay true to his philosophy of life in being daringly honest, intellectually curious and devoted to educating the future generation. I am also very grateful to my mother who despite very limited resources instilled strict discipline, and with a steady hand helped all of us “stay the course”. To all my brothers, Aravinthan, Arivu, Iniyan and Alagan, for sharing all the joys and grief in life together – I am infinitely stronger because of you.

I have been blessed to attend intellectual powerhouses since childhood - Jaffna Hindu Primary School and College, University of Peradeniya, United Nations University and Purdue University. In all these wonderful institutions, I have benefitted tremendously from my mentors and peers. Many of those relationships are close to my heart and will remain so for the rest of my life. Several colleagues I met during my time at Purdue made this long and often isolated Ph.D. experience a cherished one. I would like to especially thank all my Johnson lab mates and many other friends in both at IE and Krannert. I look forward to many more years of friendship with you.

My immense gratitude goes out to my wife, Azrah Anparasan, who had been with me almost all my adult life, for all these cherished years together. She stood like a rock when the days were short and nights were long during these laborious PhDs. I hope our son, Aanma Rumi Mahalingam, will be able to look up to his parents and be proud of the journey we took together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Technological systems contain complex elements and processes with a diverse set of agents and problem-solving arrangements. They often interact with and influence multi-lateral stakeholders with varying interests and incentives. Recent technological developments and engineering advancements such as digital marketplaces and high-tech networks create both new challenges and opportunities to understand further about effective mechanism designs. This dissertation attempts to answer corporate-level mechanism design issues in two different technological systems: high-tech biopharmaceutical networks and the online peer-to-peer lending industry.

The first part of the dissertation focuses on identifying the emergence and evolution of near decomposable systems in interorganizational relationships. To do so, first I conceptually discuss how near decomposable systems can emerge in interfirm relationships. Second, leveraging advancements in network science, I empirically analyze a detailed biopharmaceutical alliance data set and find that strategic alliance networks of biopharmaceutical firms exhibit near decomposable characteristics. I identify an emerging evolutionary pattern with smaller networks of subcommunities organizing hierarchically over time into a larger network structure, with the subcommunities generally exhibiting local clustering. A salient finding, compared to previous studies in the field of strategic management, is the identification of nested clusters formed in hierarchical fashion within this interfirm network. I find the potential for simultaneous evolutionary processes to be in play in various subnetworks within the overall industry-level network. The accrual of local changes impacting the structural processes of the subnetworks slowly diffuses to the larger, less integrated modules of the network. Finally, with the help of a simulation model, I identify how fitness heterogeneity among firms, fitness heterogeneity among partnerships and the rate of growth of partnerships impact the emergence of near decomposability in varying degrees.

The second study focuses on understanding an important market access control mechanism: platform owners granting priority access to a subset of supply-side complementors to grow the marketplace and remove potential demand-side bottlenecks. Platform governance mechanisms, such as market access control, help to align all market players towards a specific value proposition. I study the interplay between priority access and the variation in expertise of the complementors. Leveraging a randomized priority access given to expert institutional investors in the online peer-

to-peer lending industry, I show that it creates negative spillover effects on the performance of crowd retail investors. I provide evidence in support of two mechanisms in driving the impact of priority access, the intensity of priority access and cream skimming by institutional complementors, on the retail crowd market. Again using simulation to extend the analysis, I find that the brunt of negative impacts is likely borne by more risk-averse retail investors.

1. INTRODUCTION

1.1 Technological Systems and Mechanism Design

Technological systems contain complex elements and processes with a diverse set of problem-solving arrangements (Hughes, 1987; Salvendy, 2001; Simon, 1962; Simon, 2000). They often deal with multi-lateral stakeholders with varying interests and incentives (Adner, 2017; Baldwin *et al.*, 2000; Tiwana, 2013). However, all these components need to be in some form of alignment in their objectives, division of responsibilities and operational procedures for the system to function effectively (Eisenhardt, 1985; Kretschmer *et al.*, 2020; Perrow, 2011). To do so, scholars have noted the importance of uncovering the properties that could explain the tendency of varying complex systems with many autonomous components to behave according to a few selective and emergent rules (Amaral & Uzzi, 2007; Barabási, 2016; Strogatz, 2004). This is particularly important when technological systems can be partially or fully designed by engineers and managers (Buede & Miller, 2016; Roth, 2018).

Mechanism design methods provide a structure wherein design of economic mechanisms and incentives can be done with the systemic outlook towards achieving desired objectives (Mookherjee, 2006; Myerson, 1989; Roth & Peranson, 1999; Varian, 1995). It considers "how to implement good system-wide solutions to problems that involve multiple self-interested agents, each with private information about their preferences" (Parkes, 2001). In technological systems, the mechanism design framework helps to identify the collective nature of decision problems such as kidney allocation, school choice or task allocation within teams (Abdulkadiroğlu & Sönmez, 2003; Ashlagi & Roth, 2012; Roth *et al.*, 2004; Su & Zenios, 2006). It specifies then how to evaluate the process outcomes. For instance, it allows decision making in the system level to clarify the priorities such as efficiency tradeoffs, profit maximization considerations and social welfare implications. It will also evaluate the resources the agents possess such as private information and identify channels in which they can be transmitted and used in decision making (Roth, 2018; Varian, 1995).

Recent technological developments and engineering advancements such as digital marketplaces and high-tech networks create both new challenges and opportunities to understand further about effective mechanism design methods. First, mechanism design questions that once were part of operations-focused decisions have been elevated to the business strategy level by the possibilities of solutions enabled by large-scale data. For example, Uber's demand- and supply-focused surge pricing goes beyond being a mere revenue and price optimization to a core challenge of corporate strategy with implications on their business model. Similarly, if large interfirm networks with varying moving parts could be explained by a relatively small set of network scientific characteristics, it creates contingencies for firms in their corporate-level planning and strategic considerations.

Second, even as mechanism design has transformed business for several decades thanks to contributions from disciplines such as engineering, economics and computer science, scholars focused on strategy and entrepreneurship have paid relatively limited attention.[1] Traditionally, firms devised strategies to succeed in markets by building some distinct and sustainable competitive advantage, then focusing on the organizational design aspects needed to realize it. Now, the digital transformation and high-tech networks provide an opportunity for engineers and managers to design and deploy entire marketplaces, orchestrating the participation of appropriate market players and other complementary actors.

Third, when market design is a possibility, it leads to concerns about relevant incentives and other fairness considerations across autonomous firms and other agents participating in them. For example, if market designers use governance tools available to them in digital platforms and high-tech networks, they may tip the balance in favor of some players depending on their level of expertise, potentially raising anti-trust and other regulatory concerns.

1.2 Summary of the Dissertation

This dissertation attempts to answer corporate-level mechanism design issues in two different technological systems: high-tech biopharmaceutical networks and online peer-to-peer lending industry. In the high-tech biopharmaceutical networks, I use network scientific developments to show the emergence of near decomposable properties in the interactions among firms. In the online peer-to-peer lending industry, I show how an empirical mechanism design perspective helps us to understand the dynamics of decision making, incentives and information in platforms.

[1] With the general focus on identifying the factors that create heterogeneity in organizational performance, issues such as kidney allocation or task allocation within teams are not the ones to capture the attention of strategy and entrepreneurship scholarship.

1.2.1 Nearly Decomposable Systems in Interfirm Alliance Networks

The purpose of this chapter is to identify emergence and evolution of near decomposable systems in interorganizational relationships. To do so, first I conceptually discuss how near decomposable systems can emerge in interfirm relationships. Second, leveraging advancements in network science, I empirically analyze a detailed biopharmaceutical alliance data set and find that strategic alliance networks of biopharmaceutical firms exhibit near decomposable characteristics. I identify an emerging evolutionary pattern with smaller networks of subcommunities organizing hierarchically over time into a larger network structure, with the subcommunities generally exhibiting local-level clustering. A salient finding, compared to previous studies in the field of strategic management, is the identification of nested clusters formed in hierarchical fashion within this interfirm network. I find the potential for simultaneous evolutionary processes to be in play in various subnetworks within the overall industry-level network. The accrual of local changes impacting the structural processes of the subnetworks slowly diffuses to the larger, less-integrated modules of the network. Finally, with the help of a simulation model, I identify how fitness heterogeneity among firms, fitness heterogeneity among partnerships and rate of growth of partnerships impact the emergence of near decomposability in varying degrees.

1.2.2 Implications of Priority Access in Markets with Experts: Evidence from Online Marketplace Lending

Platform governance mechanisms such as market access control help to align all market players towards a specific value proposition. This chapter focuses on understanding an important market access control mechanism: platform owners granting priority access to a subset of supply-side complementors to grow the marketplace and remove potential demand-side bottlenecks. I study the interplay between priority access and variability in the expertise of the complementors. Leveraging a randomized priority access given to expert institutional investors in the online peer-to-peer lending industry, I show that it creates negative spillover effects on the performance of crowd retail investors. In this study, I exploit the October, 2012 decision of the largest online peer-to-peer lending platform, Lending Club, to create a whole loans market and implement a randomized process of allocating loans to institutional investors. Loans assigned to the institutional investors would be funded wholly by one investor. Before the implementation of this market, institutional investors that wanted to participate had to do so in the fractional market where loans are funded by multiple investors. This priority access to institutional expert investors in the whole loan market, and the varying levels of access observed over time, offers a unique opportunity to capture the negative spillover effects of this allocation on the performance among the retail investors. I provide evidence in support of two mechanisms driving the impact of priority access on the retail crowd market: the intensity of priority access and cream skimming by institutional complementors. Finally, using data-integrated simulation, I conclude that more risk-averse fractional investors likely are more strongly impacted by this effect.

1.3 Contributions

1.3.1 Research Contributions

This dissertation makes a number of contributions to the interface of engineering and strategic management scholarship. First, recent advancements in methodologies such as network science tools and large scale data availability have transformed the way traditional business organizations have operated. One such distinction is the sharp change in the way top-level corporate strategies are planned and executed within the firms. Corporate managers are starting to not merely rely on conceptual frameworks such as simple cost-benefit analysis, Michael Porter's Five Forces for industry analysis and value-based mapping. They are shifting to more nuanced data-integrated strategic management approaches. Traditionally the contributions from industrial engineering thinking to the management of firms were largely limited to the operations level. This is well reflected by the expansive operations research work focusing on business problems and the operations management scholarship in business schools. However, new methodological capabilities and the availability of data enable the use of systems thinking from engineering in multi-stakeholder environments. This dissertation demonstrates how strategic thinking could be developed and possibly used in two technological systems.

Second, a well demonstrated benefit of network science and other related methodologies is their ability to identify small sets of drivers causing large systemic level changes and leave out the other noisy elements. Applying this idea to the high-tech biopharmaceutical alliance network, I demonstrate how structural properties such as near decomposability will affect the firm- and industry-level decision making on innovation. In doing so, I show how strategy questions could benefit otherwise intellectually distant methodologies. It also elucidates how, if certain structural patterns create conducive environments, they might be considered in the formation of appropriate industries and firms.

Third, an empirically-driven mechanism design approach facilitates the development of technological systems where incentives, information and decision-making can be aligned towards desired goals. However, in complex systems it is vital to see how the interaction among different sets of agents may lead to unintended consequences hindering the incentive alignments. Exploitation of quasi-natural experimental opportunities with econometric methods provides complex systems scholarship a practical pathway to use empirical mechanism designs. For example, in my dissertation, I demonstrate that seemingly innocuous mechanism design choices made by the platform owner related to lender side expansion can create incentive conflicts and negative spillover effects for the retail investors even when the loan allocation process is randomized.

1.3.2 Practical Implications

This dissertation has several implications for practice. First, my findings provide insights regarding how to think about design and development of incentives in networked innovation environments. When structural properties can be realized in industry-level networks such as high-tech biopharmaceutical partnerships, it draws attention to both engineers and policy makers on how to orchestrate them to achieve desirable outcomes for the society. For example, during COVID-19, many governments have aggressively used war-time and other emergency powers to foster partnerships among otherwise competitors to hasten the discovery, development and production of vaccines. The U.S. Government actively worked with and mandated a partnership between Johnson & Johnson and Merck to fast-track vaccine manufacturing. Similarly, nonprofit initiatives such as a Gates Foundation initiative to get the University of Oxford to partner with AstraZeneca has spurred a successful collaboration in a short time. I hope that the theoretical and empirical findings in the dissertation provide some ideas for new innovative pathways in complex systems.

Second, the digital transformation we see in the economy today has enabled entire marketplaces to be developed applying "engineering thinking" to economics and strategy (Roth, 2018). Complex systems design thinking allows the designers to consider the importance of governance mechanisms such as market access control to get a holistic picture about the marketplaces. Platform owners can use access controls to grant, restrict and revoke market access to a selective number of complementary players. Practitioners will benefit from carefully assessing the pros and cons of available governance to effectively manage their platforms.

Third, as these high-tech networks and digital marketplaces open up new possibilities, they also create several public policy and anti-trust concerns. Scale-free and hierarchy properties in high-tech networks, and intermediated marketplaces with their congregation of experts and non-experts, create challenges with regard to the level playing field and concerns of unequal access. The Stigler Committee on Digital Platforms (2019) articulated this issue as follows: "the proposals were reactions to the perceived threat posed by digital platforms, with little to no analysis of the underlying root problems, let alone a link between market failures and remedies." Current policy infrastructure often lacks the toolkits required to understand and respond to the emerging challenges posed by these novel forms of organizations. For example, status-quo anti-trust mechanisms suggest using price theoretic mechanisms for regulating large firms. But how do we handle organizations such as digital platforms who appear to provide substantial consumer surplus at the lowest cost, yet also happen to kill their competitors and create monopolistic markets? I suggest that mechanism design solutions facilitated by big data may help us to understand the interactive effects among multiple players in a fine-grained manner, leading to more nuanced understanding of the underlying value propositions.

1.4 References

Abdulkadiroğlu A, Sönmez T. 2003. School Choice: A Mechanism Design Approach. *American economic review* **93**(3): 729-747.

Adner R. 2017. Ecosystem as Structure: An Actionable Construct for Strategy. *Journal of Management* **43**(1): 39-58.

Amaral LAN, Uzzi B. 2007. Complex Systems—a New Paradigm for the Integrative Study of Management, Physical, and Technological Systems. *Management science* **53**(7): 1033-1035.

Ashlagi I, Roth AE. 2012. New Challenges in Multihospital Kidney Exchange. *American Economic Review* **102**(3): 354-359.

Baldwin CY, Clark KB, Clark KB. 2000. *Design Rules: The Power of Modularity*. MIT press.

Barabási A-L. 2016. *Network Science*. Cambridge University Press.

Buede DM, Miller WD. 2016. The Engineering Design of Systems: Models and Methods.

Eisenhardt KM. 1985. Control: Organizational and Economic Approaches. *Management science* **31**(2): 134-149.

Hughes TP. 1987. The Evolution of Large Technological Systems. *The social construction of technological systems: New directions in the sociology and history of technology* **82**.

Kretschmer T, Leiponen A, Schilling M, Vasudeva G. 2020. Platform Ecosystems as Meta-Organizations: Implications for Platform Strategies. *Strategic Management Journal*.

Mookherjee D. 2006. Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective. *Journal of Economic Literature* **44**(2): 367-390.

Myerson RB. 1989. Mechanism Design. In *Allocation, Information and Markets.* Springer.

Parkes DC. 2001. *Iterative Combinatorial Auctions: Achieving Economic and Computational Efficiency*. University of Pennsylvania Philadelphia, PA.

Perrow C. 2011. *Normal Accidents: Living with High Risk Technologies-Updated Edition*. Princeton university press.

Roth AE. 2018. Marketplaces, Markets, and Market Design. *American Economic Review* **108**(7): 1609-1658.

Roth AE, Peranson E. 1999. The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design. *American economic review* **89**(4): 748-780.

Roth AE, Sönmez T, Ünver MU. 2004. Kidney Exchange. *The Quarterly journal of economics* **119**(2): 457-488.

Salvendy G. 2001. *Handbook of Industrial Engineering: Technology and Operations Management*. John Wiley & Sons.

Simon H. 1962. A. 1962. The Architecture of Complexity. *Proceedings of the American Philosophical Society* **106**(6): 467-482.

Simon HA. 2000. Can There Be a Science of Complex Systems. *Unifying themes in complex systems* **1**: 3-14.

Strogatz S. 2004. *Sync: The Emerging Science of Spontaneous Order*. Penguin UK.

Su X, Zenios SA. 2006. Recipient Choice Can Address the Efficiency-Equity Trade-Off in Kidney Transplantation: A Mechanism Design Model. *Management science* **52**(11): 1647-1660.

Tiwana A. 2013. *Platform Ecosystems: Aligning Architecture, Governance, and Strategy*. Newnes.

Varian HR. Year. Economic Mechanism Design for Computerized Agents. In Proceedings of the USENIX workshop on Electronic Commerce.

2. NEARLY DECOMPOSABLE SYSTEMS IN INTERFIRM ALLIANCE NETWORKS

2.1 Introduction

Scholars have long emphasized the importance of near decomposable systems in engineering and management inquiries (Egidi & Marengo, 2004; Gavetti *et al.*, 2005; Simon, 1962). With its "boxes-within-boxes arrangement of subsystems and sub-subsystems," near decomposability provides internal steady states within a larger system, facilitates fast-tracked evolutionary processes and offers relatively simplified interaction rules among myriad actors within hierarchies (Simon, 2000). Much of the research in this area has focused on how near decomposable intraorganizational interactions of subunits or managers work and, subsequently, how they bring performance and evolutionary implications for the respective firms (Ethiraj & Levinthal, 2004; Frenken *et al.*, 1999; Levinthal & Workiewicz, 2018; Simon, 2002).

Although Herbert Simon has indicated that near decomposability is "far from rare" in broader economic systems beyond formal organizations (Simon, 1996), management researchers have paid limited attention to nearly, but not fully, decomposable interorganizational relationships. A substantial body of strategic management research elucidates the importance of structural position of a firm within interorganizational networks (Ahuja *et al.*, 2012; Gulati, 1999; Schilling & Phelps, 2007; Stuart, 1998; Uzzi, 1997). This structural focus facilitates firms to "anticipate properly the complex chain of contingencies" (Granovetter, 1985) and identify appropriate strategies in networked environments (Borgatti & Foster, 2003; Uzzi, 1996). Identifying whether interfirm networks have near decomposable characteristics will allow us to decode the structural patterns of complex systems. Herbert Simon articulates the usefulness of uncovering near decomposability in systems as "allow[ing] us to factor the system, so that we do not have to deal with all of its complexity at once." (Simon, 2000).

Uncovering the existence of a structural, interorganizational pattern of clustering with near decomposable properties is important for three reasons. First, at the core, strategic-minded scholars are interested in the heterogeneity of firms (Barney, 1996; Makadok, 2003; Rumelt *et al.*, 1994). Beyond the adjacent neighborhood measures, structural properties pertaining to overall networks appear to matter in many performance measures (Borgatti & Halgin, 2011; Powell *et al.*, 2005; Schilling & Phelps, 2007). Stated differently, large scale network-wide measures, often at the industry level, can serve as a source of firm heterogeneity, and “network structure reflects much about the functioning of organizations and, possibly, their coordination failures or achievements” (Salancik, 1995). So if near decomposability exists in interorganizational relationships, they can likely assist us in understanding the antecedents of individual firm performance measures. Second, patterns formed by an internal congregation of like-minded actors in a network are not static. Firms often dynamically evaluate network implications for their value creation process. Zaheer and Soda (2009) shows how network actors seek value by creating ties disconnected from a dense and stable network; potential for structural and other network opportunities drives the evolution. Conversely, disconnected nodes seek out access to denser networks for both informational and control reasons (Burt, 2005, 2009; Gulati *et al.*, 2000). It creates an evolving equilibrium over time where network patterns prompt deployment of myriad network strategies by the actors, which in turn lead to changes in the structural properties of the overall network (Borgatti & Foster, 2003; Walker *et al.*, 1997). If the dynamics among actors lead to a nearly decomposable system, then many rules from the architecture of complexity would apply (Ethiraj & Levinthal, 2004; Simon, 1962). Hence a careful analysis of structure helps us to discover the nature of industry-level partnerships and understand exactly know what is going on at the systemic level. Third, analyzing the near decomposable network patterns helps strategists to identify the latest trends in relevant industries. This bird’s-eye view to identify generic features, large-scale organizing principles of the networks, and the drivers of change can potentially complement the ego-centric network analysis quite often employed by the strategy literature (Schilling, 2009; Uzzi & Spiro, 2005).

The purpose of this paper is to identify emergence and evolution of near decomposable systems in interorganizational relationships and the potential drivers of such characteristics. To do so, first we conceptually establish how near decomposable systems emerge in interfirm relationships. Second, leveraging advancements in network science, we empirically analyze a widely-studied biopharmaceutical alliance data set (see Schilling (2009) for a review) and find that strategic alliance networks of biopharmaceutical firms exhibit near decomposable characteristics. The accrual of local changes impacting the structural processes of the subnetworks slowly diffuses to the larger, less-integrated modules of the network. We further expand this notion, with the help of a simulation model, to identify the contingencies under which near decomposable features emerge. Using both analyses, we hope to understand the emergence of near decomposability in real world alliance networks, a manifestation of inter-organizational networks and study underlying characteristics that contribute to the emergence of it.

The analysis in this paper are two-fold. First, we utilize an existing alliance database to study inter-firm alliances and the emergence of near decomposability. We utilize the heavily-studied Recombinant Capital (RECAP) biopharmaceutical alliances database. The RECAP database provides a broad coverage and is generally a well representative sample of the wide range of biopharmaceutical alliances (Adegbesan & Higgins, 2011; Schilling, 2009). We identify an emerging evolutionary pattern of smaller networks of subcommunities organizing hierarchically over time into a larger network structure, with the subcommunities generally exhibiting local-level clustering. A salient finding, compared to previous studies in the literature is the identification of nested clusters formed in hierarchical fashion within interfirm networks. We find the potential for simultaneous evolutionary processes to be in play in various subnetworks within the overall industry-level network. Second, we develop a simulation model that tests for various scenarios that impact the generation of alliances in the real world.

For the first part, we utilize the heavily-studied Recombinant Capital (RECAP) bio pharmaceutical alliances database. The RECAP database provides a broad coverage and is generally a well representative sample of the wide range of biopharmaceutical alliances (Adegbesan & Higgins, 2011; Schilling, 2009). For the second part, we develop a simulation model that captures three characteristics that impact inter-firm alliance generation: firm-level characteristics (node fitness probability), the dyadic relationship between firms (edge fitness probability), and the network-wide propensity to create alliances (rate of growth of partnerships). Based on the simulation in this research, we observe that even though firm-level characteristics does have an impact on the emergence of scale free property in alliance networks, its variation has limited consequential impact on the emergence of hierarchical property. Hence, the overall impact on the emergence of near decomposability due to variation in firm-level characteristics is consistent. However, the varying degree of dyadic relationship (edge fitness probability) does not have a significant impact on scale free property and hierarchical structure. We observe that the preexisting relationship between firms has a similar output. The network-wide propensity to create alliances (rate of growth of partnerships), has a negative influence on the emergence of near decomposability in network structures. Further in depth analysis and detailed modeling that captures more granular firm-level, dyadic relationship level, macro-level characteristics and other inter-firm characteristics of alliance networks would be necessary to understand contingencies that may affect these results.

This paper makes a number of contributions to the literature. First, we expand the notion of near decomposability beyond the intra-organizational context. To the best of our knowledge, no other study has used either the conceptualization or data-integrated empirical network analysis in identifying the underlying near decomposable structure in an industry-level, interfirm strategic network. By providing a theoretical foundation and appropriate stylized analysis, we hope this line of research will open up new avenues for scholars to further investigate potential mechanisms originating from near decomposability on a wide variety of performance and evolutionary outcome measures. Second, it introduces new analytical developments and techniques in network science to the management literature, where it can help us understand relevant strategic inquiries on how the patterns of interfirm clustering emerge. The literature on alliances and networks has long recognized the importance of firms interacting with each other in the pursuit of creating and capturing value (Ahuja *et al.*, 2012; Gulati, 1998; Gulati *et al.*, 2000; Mesquita *et al.*, 2017). These opportunities are often generated and sustained by the ability to have distinctive heterogeneous advantage in the networks (Baum *et al.*, 2003; Borgatti & Foster, 2003; Rowley *et al.*, 2000; Schilling & Phelps, 2007; Uzzi & Spiro, 2005); we provide an analysis of how heterogeneity develops via evolutionary processes and forms hierarchical systems. Finally, a growing stream of research takes a structural approach to network studies. This approach suggests that, in networks, "beneath the complexity of social relations there are enduring patterns of 'connectivity and cleavage' [..] that, once revealed, can help explain outcomes at different levels" (Kilduff & Brass, 2010; Wellman & Berkowitz, 1988). The structuralist strand of scholarship provides an avenue for linking micro-level nodal concerns to macro-level observations and, also conversely, macro-level contingencies to micro-level changes. By providing a nested hierarchical linkage to local clustering, we contribute to this stream of literature.

The rest of this paper is organized as follows. Section 2 gives a detailed description of interfirm alliances and the characteristics of near decomposable systems. Section 3 covers the empirical analysis using RECAP biopharmaceutical alliance data. In section 4, we explain the simulation model and report results of the carious scenario analyses that we conduct to tease out the firm-level, dyadic-relationship level and macro-level characteristics that impact the emergence of near decomposability in alliance networks. Section 5 summarizes the results observed in this paper and highlights future research directions.

2.2 Industry-Level Inter-Firm Network and Near Decomposability

2.2.1 Complex Systems Perspective and Near Decomposability

We adopt Simon's conception of complex systems as "one made up of a large number of parts that interact in a non-simple way... given the properties of the parts and the laws of their interaction, it is not a trivial matter to infer the properties of the whole" (Simon, 1962). The implications of complexity in a near decomposable system stem from how the interrelated subsystems interact each other, and the magnitude and frequency of their relationships (Ethiraj & Levinthal, 2004; Simon, 2000). Simon (1962) summarizes the nuances of analyzing complex systems by noting that "an in-principle reductionist may be at the same time a pragmatic holist."

The pursuit of strategic alliances by a firm is a conscious decision to collaborate with another, believing that partnering is likely to yield a better outcome than internally executing the same task themselves. So in an important pragmatic sense, the interdependent interactions among firms in an alliance network mean that, at the aggregate level, "the whole is more than the sum of the parts." Due to the intensive nature of competition and the need to be on the cutting edge of innovation, biopharmaceutical firms closely monitor developments in the innovation and commercialization space of their market. Prior studies suggest that technologically intensive strategic alliances are resource-interdependent (Dhanaraj & Parkhe, 2006; Dyer & Singh, 1998; Gulati *et al.*, 2012). They seek, synthesize and share information in partnership networks (Gulati, 1998; Koza & Lewin, 1998). Alliance partners utilize a recombinatory innovation process to resolve innovation problems (Gilsing *et al.*, 2008; Schilling & Phelps, 2007). Though these studies indicate "nonsimple" interactions among alliance partners, they also show that, under appropriate contingencies, we can tease out the broader principles of partnerships. This tension provides a complex systems outlook to alliance networks.

The near decomposable property in complex systems, if it exists, provides a pathway to handle substantial complexities in the system without having to deal with them all at once. The following comment is representative of Simonian thinking on the usefulness of near decomposability in complex systems (Simon, 2000):

"Having determined the behavior of subunits at one level, we can replace the details of these subunits by a small number of aggregate parameters, and use these to represent the system at the next level above. Or, looking from the top down, we can say that the behavior of the units at any given level does not depend on the detail of structure at the next level below, but only upon the steady state behavior, in which the detail can be replaced by a few aggregated parameters."

Prior research specifically indicates the value of discernible, yet connected, couplings where the order emerges as most activities are completed at the local level, e.g., within knowledge clusters (Yayavaram & Ahuja, 2008), subproblems (Nickerson & Zenger, 2004), and divided managerial labor (Frenken *et al.*, 1999; Levinthal & Workiewicz, 2018). Yet this order is consolidated at the complex systemic level due to the fact that "interactions among knowledge sets within subproblems are greater than among subproblems" (Nickerson & Zenger, 2004). This process helps actors in the complex systems to be able to, if needed, to temporarily neglect broader coordination issues and focus on the specific subsystemic issues (Baumann, 2015). Absence of this coordination can lead systemic equilibrium to be stuck in suboptimal local peaks in the fitness landscape (Rivkin & Siggelkow, 2003). Studies from a more behavioral perspective suggest that preemptively acknowledging the need for an appropriate level of coordination can assist boundedly rational agents in complex systems to tackle issues such as myopia of learning (Levinthal & March, 1993), to identify the right level of landscape search (Gavetti *et al.*, 2005) and to use their limited capacity to more effectively process necessary information (Reinstaller, 2007).

Near decomposability, from the Simonian perspective, suggests two additional, and often neglected, properties that facilitate systemic functionalities: the nested nature of subsystems (Levinthal & Workiewicz, 2018)[2] and the general horizontal similarities between those subsystems. Nested hierarchy can be crudely described as analogous to nested Russian dolls or Chinese boxes. While the hierarchy in the dolls and boxes imply a complete ordering, multiple interlinked sets – several cascaded dolls and boxes – suggest a partial order similar to a tree (Simon, 1977; Wu,

[2] Levinthal and Workiewicz (2018) summarizes the first part of the concern well here in an NK modeling context: "[..]the difference between modularity and near decomposability has not been fully addressed [..]. Scholars have used the term modularity to describe patterns of interactions between individual elements of a system in which most of the interactions occur within the modules (within interactions), with only relatively few interactions between elements lying in different modules (between interactions). The discussion of NK models often uses the terms "modularity" and "near decomposability" interchangeably [..]. However, these NK models, with "block-diagonal" interactions, capture only a subset

2013). This partial order in the form of encapsulating network hierarchy facilitates a vertical coupling across different tiers in the hierarchical system and "permits the stable subassemblies to be treated as simple givens, whose dynamic behavior is irrelevant to assembling the larger structures, only their equilibrium properties affecting the system behavior at the higher levels" (Simon, 1977). However, horizontal similarities among those subsystems – for example, if the third doll in a set of dolls is similar to the third doll in another set – provide a larger framework for the system to rely on. It allows each subsystem to simultaneously and independently execute the relevant tasks, needing only the necessary inputs from the rest of the system and producing output as needed after the subroutine is performed within the subsystem. This form of loose horizontal coupling allows systems to modify internal subroutines without disturbing subsystemic inputs and outputs.

2.3 Near Decomposability and Interfirm Alliances

In the realm of complex inter-organizational alliance networks, studies already show the importance of being at a relevant network position. Research has looked into how firms continue to pursue structurally friendly locations (Gulati, 1998; Gulati *et al.*, 2000; Powell *et al.*, 1996; Zaheer & Bell, 2005) and, conversely, how the lack thereof can be detrimental for their performance and survival (Dyer & Singh, 1998; Koza & Lewin, 1998). Other inquiries of a structural nature include studies of the aggregate number of ties (Ahuja, 2000; Shan *et al.*, 1994), density of the ego network (Hite & Hesterly, 2001; Rowley *et al.*, 2000), small world properties (Baum *et al.*, 2003; Uzzi & Spiro, 2005) depth (Schilling & Phelps, 2007) and relative standing (Baum *et al.*, 2000; Ozmel & Guler, 2015). Network measures mostly originating from the field of sociology have improved the sophistication of empirical analysis in an interdisciplinary literature on network analysis in the field of strategic management (Ahuja *et al.*, 2012; Burt, 2001, 2009; Coleman, 1988; Coleman, 1994; Zaheer & Soda, 2009). However, conceptual articles such as Ahuja *et al.* (2012) and clinical cases like Gomes-Casseres (1994) notwithstanding, researchers have devoted limited inquiry to understanding how aggregate systemic patterns emerge in industry-level, interorganizational networks.[3] Network patterns originate from the inherent heterogeneity of how nodes interact with their fellow network actors. They start to show clustering

[3] Salancik (1995)articulated this general tendency in organizational network studies as the penchant to "focus on the trees rather than the forest."

patterns – "the emergence of interconnected subgroups, or network partitions or cliques, suggests that the network is being differentiated into a variety of distinct subnetworks or communities" (Ahuja *et al.*, 2012). Nearly decomposable systemic patterns in interorganizational relationships describe these complex heterogeneous nodal interactions in networks rather efficiently, as facilitated by all three mechanisms associated with near decomposable systems – (i) interactions within subsystems are greater than among subsystems, hence a higher degree of clustering is observed and subsystems exhibit (ii) a nested, hierarchical nature and (iii) horizontal similarities among subsystems on the same tier of hierarchy. We discuss their implications below individually.

In alliance partnerships, interactions within subnetworks tend to be greater than among subnetworks for informational reasons. A structural outlook regarding information transmission suggests that it "go[es] beyond the immediate ties of firms and emphasize the informational value of the structural position these partners occupy in the network. Information travels not only through proximate ties in networks, but through the structure of the network itself" (Gulati, 1998).[4] Clustered subnetworks increase both information transmission and processing capacity (Schilling, 2009; Yayavaram & Ahuja, 2008). Within subnetworks often a dense, intensively interactive and highly frequent participatory process unfolds.

2.4 Scale Free Property, Hierarchical Networks and Near Decomposability

A large number of real-world networks are, in some variation, found to have what is known in the network science community as hierarchical network modeling properties (Barabási, 2016; Barabási *et al.*, 2003; Boccaletti *et al.*, 2006; Clauset *et al.*, 2004; Girvan & Newman, 2002; Newman & Girvan, 2004; Ravasz & Barabási, 2003; Ravasz *et al.*, 2002). Some prominent examples include internet links, mobile phone calls, collaborations between scientists, actors in motion pictures, scholarly citations and metabolic networks. However, the nuances of individual hierarchical network properties differ for each of these examples in interesting ways (refer to Barabási (2016) for further discussion). Barabási *et al.* (2003) offers one simple interpretation on how hierarchical networks form: "[W]e should not think of... coexistence of relatively independent groups of nodes. Instead, we have many small clusters that are densely interconnected. These

[4] Gulati (1998) distinguished it from a more relational approach that emphasizes the role of direct cohesive ties in promoting a "shared understanding of the utility of certain behavior as a result of discussing opinions in strong, socializing relations".

combine to form larger, but less cohesive groups, which combine again to form even larger and even less interconnected clusters. This self-similar nesting of different groups or modules into each other forces a strict fine structure on real networks."

The emergence of near decomposability can be explained using the scale free property and clustering in networks. The scale free property refers to the behavior of a network's degree distribution and its adherence to a power law distribution. The clustering coefficient explains to the extent to which nodes are clustered to create tightly knit triplet groups. The emergence of near decomposability is where the clustering coefficient is a decreasing function of the degree of a node compared to a random network. In the following sections each of these characteristics, their technical definitions and how they will be used in the analysis for the purpose of this research will be explained in further detail. Figure 2.1 shows the illustration of a hierarchical network.

2.4.1 Scale-Free Property.

If a node in a network has links to k other nodes, then we say that it has degree k; the average degree of all nodes in a network is denoted $\bar{k}$. The network satisfies the scale-free property if the frequency distribution of node degrees approximates a power-law distribution, i.e., if the probability of a randomly-selected node having degree k is distributed as $P(k) \sim k^{-\gamma}$ for some constant γ. This indicates that the network has a small number of highly-connected nodes, such that the number of nodes with a degree far greater than the average degree of the network (i.e., k ? $\bar{k}$) with is significantly greater than the number present in a random graph. One of the key mechanisms proposed in the network science literature for producing a scale-free network is "preferential attachment," the notion that nodes are more likely to form attachments with other nodes that are already highly-connected (Ravasz *et al.*, 2002). Albert and Barabási (2002) provides an example in internet, "a webpage will more likely include hyperlinks to popular documents with already-high degree, because such highly connected documents are easy to find and thus well-known, or a new manuscript is more likely to cite well-known and thus much-cited publications than less-cited and consequently less-known papers". It is close to the colloquial saying that "the rich get richer, and the poor get poorer." As a network evolves and forms new links over time, the probability of a link attaching to a node is assumed to be an increasing function of the node's degree in the previous time period. The functional form and strength of the preference may vary in different real-world contexts, but this concept of preferential attachment is consistent with plausible firm behavior.

2.4.2 Clustering Coefficient

The local clustering coefficient of a node is defined by $C = \frac{2n}{k(k-1)}$, where k is the node's degree and n is the number of links between those k neighbors. The average clustering coefficient of the network, $\overline{C}$, is the average of every node's local clustering coefficient. We can further define a function $C(k)$ to be the average clustering coefficient over all nodes with degree k. The normalized clustering coefficient of a node is significantly higher in real networks than random networks and the clustering coefficient of real networks is independent of the number of nodes in the network.

2.4.3 Emergence of Near Decomposability.

In a hierarchical network, the clustering coefficient of a node $C(k)$ is a decreasing function of degree k, and it follows the relationship $C(k) \sim k^{-\beta}$ for some β with a value that is typically close to 1. This indicates that less-connected nodes are highly clustered together; by contrast, well-connected nodes with high degrees serve as hubs of communication for the network, generally linking to less well-connected neighbors. Combined with the scale-free property, this produces hierarchical structures wherein clusters of nodes connect to other clusters through their hubs, and clusters of clusters connect to each other, in a process analogous to galaxies of [clustered] stars forming clusters of galaxies under the weight of gravity.

Thus in this research, we test for scale free property of a network and the scaling behavior of local clustering with degree (hierarchy) to describe the emergence of near decomposability in the overall network. Following these arguments, we hypothesize that in interfirm networks, smaller network subcommunities recursively organized hierarchically into a larger network in a nested fashion. These nested subcommunities emerging from the hierarchical network exhibit local-level clustering.

2.5 Empirical Analysis

2.5.1 Data

To illustrate the importance of understanding patterns of clustering and hierarchy within a strategic management context, we utilize the heavily-studied Recombinant Capital (RECAP) bio pharmaceutical alliances database. It has 12,962 instances of alliance partnerships between 5,524 firms during 1985 - 2001. Recent studies suggested the RECAP database provides the "broadest relative coverage" and "representative samples" of the wide range of biopharmaceutical alliances (Adegbesan & Higgins, 2011; Schilling, 2009). Hence, it serves well as an appropriate context for our study. Also, it is worth noting that substantial strategic inquiries coming from the Coleman-Burt tradition have used the RECAP data to study partner selection (Laursen *et al.*, 2013; Mindruta *et al.*, 2016), reputation and status (Stern *et al.*, 2014), complementarity of assets (Hess & Rothaermel, 2011) and exploration and exploitation (Yang *et al.*, 2014) – these studies provide a diverse set of examples illustrating the versatility of the database.

If firm i created an alliance with firm j at time period t, then the network created for our analysis represents this as an edge between nodes i and j. Hence, if there were multiple alliances between two firms, then we allow multiple edges between the respective nodes. [5]

2.5.2 Results

The analysis on the alliance network are explained in the following order. First, we describe the evolution of the overall composition of the types of alliances that were created and specific key characteristics about firm behavior in the overall network; second, we test for the scale free property; third, we test for the clustering coefficient and its behavior over time; then we explain the emergence of near decomposability in the network. Finally, we also include a subsection of complementary analysis, where we show additional analyses on the overall network for deeper understanding of the analyses conducted in the previous sections.

[5] We break down the network into periods of 3 years and 5 years respectively. 3 and 5-year periods are used as a sensitivity check based on the fact that the data set does not note the duration of alliances. Here we follow the approach used by Schilling (2009).

2.5.3 Overview of Alliance Networks.

There are three types of entities create partnerships with each other in the data set. They are biotech (Bio), pharmaceutical (Pharma) and universities. Figures 2.2 – 2.4 give a network-based view of the evolution of the alliances, and Table 2.1 gives a clear breakdown of how the different types of alliances evolved over 5 year consecutive periods from year 1985 to 2001. In the initial 5 year period, 1985-1989, we observe that the largest number of alliances were Bio-Pharma (56.19%) followed by Bio-University (22.02%). There were comparatively fewer players in the market and the thick edges show that a large number of alliances were made between certain established entities. However, over time the share of Bio-Pharma drops down to 35.23%, by 1997-2001, overtaken by Bio-Bio alliances (42.37%). Bio-Bio alliances is the category that rose most drastically, from a mere share of 8.57% in 19851989 period to lead the alliance market at 42.37% by 1997-2001. From general observation of the networks we see that: (1) the number of firms creating alliances has increased (2) firms tend to create fewer alliances with a given firm (thin edges) by 1997-2001, hence diversifying their portfolio of alliance relationships, compared to 1985-1989 (where we see thicker edges between firms) (3) the number of firms that are not connected to the giant component[6] is also large. These firms create a single alliance between them and are not connected to any firm that is connected to the giant component of the network (they are represented in the outer boundary of the network where they are typically connected to only one or few other nodes that are not connected to the giant component).

2.5.4 Testing for Scale-Free Property.

To test whether the data represents scale free property over time, we need to assess if the degree distribution of the network follows a power law curve. Figure 2.5 presents the number of alliances a potential firm has with other firms for consecutive 5-year time periods.

The graphs represent the cumulative degree distribution for each of the time periods considered. As the years become more recent, with a larger number of firms and far more alliances created, we see that the graph shifts to the right. The linear relationship in the data become more

[6] A giant component of a network considers the connected nodes that constitute a significant part of the complete number of nodes and edges that make up the network. In the alliance network data we find that there are a smaller number of firms unconnected to the giant component, these tend to be newer firms that are just starting to create alliance partnerships that can be considered under bio-pharmaceutical alliances.

evident with time. Even though minor deviations from linearity are observed in the data, the overall linear relationship stays intact.

To test the scale free property, we hypothesize that the linear relationship that seems to develop over time in the data follow a power law distribution. In order to test whether the degree distribution actually follows a power law distribution, we conducted the Kolmogorov Smirnov (KS) test on the data, and power law fitted distribution values. The KS Statistic in this case, will help us compare a one dimensional probability distribution (in this case the degree distribution of the alliance network), with a reference probability distribution (in this case, a power law distribution). The test compares the distance between the two distributions. The test provides the numeric scalar, alpha, which is the exponent of the fitted power-law distribution and the KS Statistic, which compares the fitness of the empirical distribution to the reference distribution. The smaller the score of the KS Statistic, denotes better fit. The p-value of the test gives an indication whether the test rejected (or vice versa) the null hypothesis that the empirical data fits a power law distribution. A P-value of <0.05 would indicate that the null hypothesis was rejected.

Table 2.2 provides the summary KS test statistics. The summary provided here is only for the giant component, since applying the KS-Statistic on the entire data set including the loosely connected nodes yields poor results, hence it is evident that only the giant component adheres to the scale free property. Based on the formal model description, beta is the power law exponent. If the data follows a power law distribution, then $2 \leq \beta \leq 3$. While it can still be a power law distribution with other exponents, empirical analyses commonly find that real-world power law distributions to have exponents between two and three. We see that over time, as the number of alliances between firms and the number of firms in the network increases, the scale free property emerges. The KS-Statistic column gives the difference between the actual distribution and the fitted power law distribution. This value needs to approach zero for perfect fit between the two distributions. The P-Value column, represents the p-value of the KS-Statistic, if this value is greater than 0.1, then we can conclude that we have sufficient confidence that the alliance data follow a power law distribution. [7]

[7] KS-Statistic was computed such that a p value <0.05 indicated rejection of the null hypothesis, indicating that the sample data followed a power law distribution. We increase the threshold to 0.1 in our case as an indication of a higher threshold.

2.5.5 Testing for Characteristics of Network Clustering.

Here, we test for the behavior of node-level clustering over time. The logarithm of local clustering of nodes are plotted against the logarithm of their degree. In Figure 2.6 this behavior is plotted for consecutive 5 year periods. The overall pattern observed in clustering is a downward slope with node degree. This behavior illustrates is a correlation between node level clustering and the degree of a node. The downward slope, while not completely aligned to a strictly linear line, the expectation is to observe whether there is correlation with degree of a node.

2.5.6 Emergence of Near Decomposability.

Following the logic articulated in the previous section, we explain the emergence of near decomposability using two characteristics in the network structure: the scale free property and hierarchical structure.

Above we observed that the alliance network, over time, displays the scale free property since the degree distribution follows a power law. The adherence of the KS statistic within required parameters is consistent with the hypothesis.

The next characteristics that determines near decomposability is the hierarchical structure of a network. There is a hierarchical nature to a network if there is a scaling behavior between local clustering and degree of a node. As the degree of a node increases, the local clustering coefficient has to decrease. On the other hand, if there was no hierarchical structure, then there would be no scaling relationship between the local clustering coefficient and degree. Again, we observe that this scaling behavior is present in the alliance network, hence we can determine with reasonable confidence that the network has a hierarchical structure (Refer Figure 2.5 and Figure 2.6).

2.5.7 Testing for 3 Year Periods.

We conduct a robustness check to determine whether the conclusions derived regarding scale free property and hierarchical structure stand true when we consider 3 year time periods. Figure 2.7 displays the behavior of degree distribution for consecutive 3 year time periods. The linear behavior that develops over time as the graphs move right from left, can be hypothesized to follow a power law distribution. Table 2.3 indicates that even though there are deviations from the power law distribution in the initial periods (in the initial periods beta does not fall between 2 and 3), the data start adhering more to power law distribution over time. The existence of limited alliances (lower number of nodes and edges) compared to a 5 year period, and hence lesser dense graphs potentially contribute to the minor deviations that are observed in the beta values over time compared to the 5 year period. Figure 2.8 shows the local clustering against node degree. Again, over time we see a clear scaling pattern where the node level clustering reduces as degree increases. These behaviors combined, suggest that by and large, the scale free property and hierarchical structure exist even when we assume 3 year time periods. However, the emergence of near decomposability is more consistent with an assumption of 5 year time periods.

2.6 Simulation Analysis

2.6.1 Introduction to the Simulation

In the empirical analysis, we identified that a hierarchical architecture emerged over time in the biopharmaceutical alliance data. This is consistent with the hypothesis that mature strategic alliance networks, such as biopharmaceutical firms exhibit near decomposability. The emergence of a hierarchical architecture in the alliance network is rooted in who connects to whom over time. Since we have a multitude of organization types-bio, pharmaceutical, universities, and non-medical- in the data, these associations can be influenced by many external organizational and macro factors. Hence, delineating what drives the emergence of hierarchy in alliance networks becomes challenging. To extract factors that drive the emergence of communities, and the hierarchical nature, we must therefore inspect a network's evolution by removing the noise from other external factors.

For the purpose of identifying factors that drive hierarchical behavior in alliance networks, we develop a robust simulation model to represent inter-firm technology partnerships. We attempt to understand under what conditions scale free property and near decomposability would emerge. Broader literature suggests that several classes of determinants that play key role in the evolutionary process of alliance partnerships. Abstracting out potential underlying real world reasons for the emerging structure provide an opportunity to infer some managerial and engineering implications based on purely the impact of these determinants on the emergence of hierarchical properties in alliance networks. To do so, we assume a fixed environment and focus on firm-specific characteristics, alliance-specific characteristics and partnership growth rate. We use the simulation model to develop a fitness landscape in interorganizational setting to help understanding the impact of three key determinants on the evolutionary processes of an alliance network under a range of scenarios. These key determinants are:

1. Fitness heterogeneity among firms,
2. Fitness heterogeneity among partnerships and,
3. Rate of growth of partnerships

The goal of this simulation is to answer the following questions:

1. Do each of the factors mentioned above contribute towards the emergence of hierarchy within an alliance network?
2. Does one factor have a greater impact in the emergence of hierarchy than another?, and
3. For each of these factors, what is the pattern in which hierarchy emerges in these networks over time?

2.6.2 The Simulation Model

The simulation is developed to observe the emergence of network patterns over time, in which alliances among firms are created and terminated under specific parametric conditions. The simulation model is designed to reproduce key characteristics observed in real world alliance formation in an inter-firm network; the existence of heterogeneous firms with different levels of attraction to form alliances with each other, the formation and termination of alliances over time and the increase or decrease of attractiveness of a firm over time as an alliance partner. In the following subsection we cover details of the simulation model: parameters and variables used in the model, the rule-based attachment and detachment functions and a summary of the model execution methodology.

2.6.3 Simulation Model Formulation

The formulation of the simulation model for an alliance network and its evolutionary characteristics are explained in this section. Representation of firms is limited to a predefined specific number, N, of nodes in the network. Initially, none of these firms have any alliances created. Hence, the state space, with N firms, is defined as the total number of firms that have the "potential to participate" in creating alliances within a specified number of time periods, T. Any firm $i \in N$ can create an alliance with any other firm $j \in N(i \neq j)$. The simulation begins with no edges (alliances) created among any of the nodes (firms). This "clean slate" approach was used so as to have a clear baseline that helps compare the results of each of the scenarios that will be tested. The model does not discriminate based on "type" of alliance; i.e., no predefined characteristic is attached to an edge before or after it is created between two nodes. Two nodes may have more than one edge linking them, since there can be multiple independent alliances between two firms, each edge created at the same time period or different time periods. Table 2.4 provides the definitions for the main parameters and variables in the model.

Next we go into details of the simulation model characteristics. The simulation model, as summarized below, comprises of the following main characterizations that play a significant role in the network evolution process: 1) defining how large the state space should be, 2) randomly assigning the starting node and edge fitness probabilities, 3) randomly picking the number of alliances to be created and terminated in each time period from predefined distributions, and 4) updating the edge fitness probabilities for the next time period.

Network Alliance Simulation Model:
Define the state space; a network with set number of firms; N
Define the time horizon; T
Assign node and edge fitness probabilities; β_i, e_{ij}
For each time period $t \in T$
    Randomly pick the total number of alliances to be created from a distribution $\emptyset$
    Randomly pick the total number of alliances to be terminated from a distribution φ
    Randomly pick the nodes between which alliances will be created
    Randomly pick the nodes between which alliances will be terminated
    Update edge fitness probabilities; e_{ij}

The State Space.

The state space is defined as a matrix of size $N * N$. Since there are no alliances between the firms at the beginning of the simulation, the state space is initially defined as a null matrix. As time progresses, at each time period $t \in T$, a number of alliances get created, and a number of alliances get terminated. This populates the state space, helping us to study the conditions under which the alliance network evolves to display various real-world network characteristics.

Node Fitness.

Extant literature discusses how firm-level characteristics affect the performance of strategic alliances and business partnerships (Ahuja, 2000; Powell *et al.*, 2005; Reuer & Ragozzino, 2006). Firm-level characteristics such as firm age, firm size, innovative experience, resource and R & D capabilities, and firm's historic alliance experiences are particularly salient in the formation and evolution of strategic alliances (Ahuja, 2000; Gulati, 1998; Gulati *et al.*, 2000; Harrigan & Newman, 1990; Powell *et al.*, 2005; Reuer & Ragozzino, 2006). This specific attribute pertaining to the "firm-level" characteristics is represented by node fitness.

The node fitness μ_i is defined as the "attractiveness" of a firm that impacts the overall likelihood of creating an alliance with another firm. For the purpose of the simulation, we assume a static node fitness, as defined at $t = 0$. The node fitness value is randomly picked from a predefined distribution. The fitness values are normalized to create node fitness probabilities. As explained below, node fitness probabilities are utilized when updating edge fitness probabilities that determine the propensity for alliance creation between two firms.

Edge Fitness.

The attributes at the alliance level between two partners such as alliance specific investments, prior ties, innovation in the partnership, frequency and historic length of alliances between the partners, and type of partnerships are particularly important for the formation and evolution of alliances (Ahuja *et al.*, 2012; Gulati, 1999; Schilling & Phelps, 2007; Stuart, 1998; Uzzi, 1997). This specific attribute pertaining to the strength of "ties" between two firms is indicated by the edge fitness.

Similar to the node fitness, at time $t = 0$ we also randomly assign edge fitness between two firms, $i \in N$ and $j \in N$ from a predefined distribution. Once this parameter is normalized across all $N * N$, we use the term "edge fitness probability" to denote the probability of an edge, in other terms the propensity of an edge to be picked among other edges as a potential alliance at time t. Since the attractiveness for a potential alliance between two firms can change over time, the edge fitness probability is updated at every time period, taking into consideration the node fitness probabilities between firms (μ_i and μ_j), the number of alliances between i and j in the previous time period $t-1$ (we assume that active partnerships play a critical role compared to "all" historical partnerships between two firms, since considering every partnership historically may result in over crowding of the network) and a constant θ that serves as a network specific parameter. The importance of considering the state space at $t-1$ when adjusting edge fitness probabilities for time t is because a greater number of alliances created between two firms may have a positive spillover effect over future alliances. We consider θ to represent an alliance network specific constraint; the impact of such a constraint and defining it accurately requires empirical analysis at a much deeper level than this thesis allows.

$$e(i,j)_t = e(i,j)_{t-1} + \theta \textit{ (number of alliances)} (i,j)_{t-1}(\mu_i + \mu_j) \qquad \text{for } t>1 \qquad (1)$$

The Attachment Function.

In the simulation, alliances are created in each time period t based on an attachment function. The number of alliances to be created in each period is randomly chosen from a specified distribution φ , where the probability distribution parameters can be adjusted. Once a number of alliances to be created is picked, the model does a random sampling of the edges to be created in the state space. The random sampling utilizes the edge fitness probability measure explained earlier. Based on the formulation of the edge fitness probability for $t>1$, agglomeration of nodes takes place over time where firms with greater number of alliances become more central in the network.

The Detachment Function.

In each time period t, once alliances are created we then proceed to determining which alliances need to be detached. The number of alliances that need to be detached in a time period is determined based a predefined probability distribution ω. In each time period, it is ensured that the maximum number of alliances that can be detached is lower than the maximum number of alliances created. This ensures that overall, there is a growth in the number of alliances created in the network. The edge fitness probabilities are utilized to randomly sample the alliances that need to be detached. Using the same edge fitness probability values for the attachment and detachment functions ensures that there is no over indexing on certain nodes, which would affect the overall evolution of the alliance network.

2.6.4 Scenario Generation

As highlighted in Gulati (1998), the main sequence of events that take place in alliance formation are: 1) the decision to enter into an alliance, 2) the choice of an appropriate partner, 3) choice of the structure of the alliance, 4) the dynamic evolution of the alliance as the relationship develops over time. In our simulation, we use a Monte Carlo simulation technique to determine how many alliances will be created in a given period of time and between which firms these alliances will be created (the decision of creating an alliance and choice of an appropriate partner). We use two firms' history of alliance partnerships to update the future probability of creating an alliance between those firms (a form of the dynamic evolution between two firms and how their relationship develops over time), and we study the growth rate of alliances (difference between how many alliances are created vs. how many are terminated) as a means of capturing the market's propensity to create alliances. Hence, our simulation focuses on three factors that are critical in alliance formation: firm-level characteristics, partnership-level characteristics and external market requirements for creating alliances and partnerships. In generating scenarios to analyze what drives hierarchy in alliance networks, these factors are brought to life by: the node fitness probability, edge fitness probability and the rate of alliance generation. This subsection covers the reasoning behind selecting these factors for scenario analysis and a description of the scenarios analyzed.

Factor 1- Node Fitness Probability:

In the simulation model, we use node fitness probability as a factor that helps emulate these firm-level characteristics that are highlighted in the extant literature. The node fitness probability is initially assigned to a firm based on a probability distribution function. Over the time horizon of the simulation. Therefore, while some firms will have a higher probability compared to others in the state space, this is assumed to be reflective of a firm's technological capability, attractiveness and maturity in the market, competitive resources offered and being a better alternative compared to other firms in the market. We assume that the node fitness probability, once defined, stays constant over the entire time horizon of the simulation. Since a goal of this simulation is to better understand the impact of the "propensity" of a firm to create alliances in the state space, and not the impact of the evolution of firm-level characteristics over time, it is justifiable to maintain a single node fitness probability for a firm over time.

In the simulation, we run scenarios for a wide range of parameter values in the distribution function. This helps to test a state space for which there is a low distinction between firms to a very high variation of firm-level characteristics. Thus, emulating a range in which we see all firms have almost same "capability and attractiveness" to create alliance to a much broader variation among firms.

Factor 2 - Edge Fitness Probability:

In this simulation, the edge fitness probability is a metric that indicates the potential dyadic relationship between two firms. Empirical evidence suggests that opportunities for collaboration and alliance formation is not merely a "search" for compatibility, but that existing relationships among alliance partners plays significant role in creating future alliances. Furthermore, "Social networks of prior ties not only influenced the creation of new ties but also affected their design, their evolutionary path, and their ultimate success" (Gulati, 1998).

In our simulation model, the edge fitness probability is assumed to capture this notion of future alliance creation based on compatibility between two firms and their history of being alliance partners. In a previous section, we covered how edge fitness is assigned based on a predefined probability distribution function and how it is updated in each time period to incorporate the prior ties between two firms.

In our simulation we explicitly consider the relational strength between two alliance partners if they have prior ties between them in time $t-1$ (see equation 1). However, the structural connections of a firm and its influence of creating alliances with a partner of a partner (assume a firm A being connected to another firm B that has a large number of alliances with a central node C, hence the propensity of A being connected to C) is not explicitly modeled in this simulation. Due to computational complexity and the need to focus on understanding the overall impact of a direct dyadic relationship over time, we restrict the scope of influence of prior ties in this simulation only to direct alliance partners.

Similar to the node fitness probability, in the simulation, we run scenarios for a wide range of parameter values in the distribution function. This helps to test a state space for a range of potential “propensity of firms to create an alliance” from a low to a very high level of attractiveness between two firms.

Factor 3 – Rate of Growth in Partnership:

In this Apart from the node fitness probability, a representation of the firm-level characteristics that affect alliance formation, and edge fitness probability, a representation of the propensity of two firms to create an alliance, we also study the impact of the rate of growth of partnerships on the evolution and formation of hierarchical structure in alliance networks. Previous literature has shown that environmental attributes such as institutional environment, business climate, growth rate and cultural factors (Hamel, 1991; Hitt *et al.*, 2004; Oxley, 1999; Park *et al.*, 2002) are critical to do understand strategic alliances. For our purposes control rest of the environmental characteristics and focus on the rate of growth in partnerships. We define rate of growth of partnerships as the net number of alliances that is formed in a given time period. Therefore, a high rate of growth of partnerships corresponds to a greater net number of alliances formed in any given time t.

Unlike the node and edge fitness probabilities, which capture intrinsic characteristics at firm and dyadic levels, the rate of growth of partnerships is used to represent the nature of the market and the overall network-wide propensity to create alliances. We use this as a metric that covers the impact of industry-wide strategic behavior that may lead to; a greater propensity to form alliances that help firms stay competitive, a greater willingness of firms to leverage technological and critical knowledge of other firms, potential market growth, increased number of firms entering the market (increased number of active participants in the state space), increased efficiencies in the market that drive greater collaboration, and a varied number of other reasons driven by macro trends (Berg & Friedman, 1978; Hagedoorn, 2002; Madhavan *et al.*, 1998).

Similar to the node and edge fitness probabilities, in the simulation, we run scenarios for a wide range of parameter values in the distribution function that defines the rate of growth. This helps to test a state space for which there is a low overall net new alliances created between firms to a very high number of net new alliances created at any given time period t. Thus, emulating a range of potential macro-economic trends that drive varying levels of net new alliances created in the network.

Scenario Generation

As depicted in Table 2.5, for the simulation, we study a total of 30 scenarios. The first 10 scenarios focus on the variation in *node fitness probability distribution*. For the purpose of generating a distribution, we use a uniform distribution model with upper and lower bound parameters. The larger the difference between the upper and lower bounds, the greater the variation in node fitness probability across the nodes in the entire state space. Scenarios 11-20 focus on the variation in *edge fitness probability distribution*. The larger the difference between the upper and lower bounds in the distribution, the greater the variation in the propensity to create alliances between the firms. The first 10 scenarios focus on the variation in *node fitness probability distribution*. Scenarios 21-30 focus on the variation in *rate of growth of partnerships*. Since the rate of growth is defined as the difference between the potential alliances that are created and those that are detached during a time period t, the growth probability distribution is shown as a difference between two uniform distributions (distribution function from which umber of alliances to be attached is picked, φ, and the distribution function from which number of alliances to be detached is picked, ω. The larger the difference between the upper bounds in each of the distributions, the greater the number of potential alliances created in the network.

Each simulation run consisted of a state space where $N = 100$. The simulation was run for a time horizon of $T = 500$. Because we use probability distributions, to ensure that the results in the analysis are not skewed, each scenario was re-run 25 times. The results reported in the next section, for each scenario, is the average based on 25 runs per scenario.

2.6.5 Simulation Results and Analysis

As highlighted previously, to test the emergence of near decomposability in the alliance network, two characteristics in the network structure needs to be established: the scale free property and hierarchical structure. In this section, we analyze results of the simulation and determine the impact of node and edge fitness probability and alliance growth on the emergence of scale free property and hierarchy in the simulated networks.

Scale Free Property.[8]

A network that shows scale free property when the degree distribution follows a power law. Therefore, as explained in the earlier section, the degree distribution of a scale free network follows, as $P(k) \sim k^{-\gamma}$ for some constant γ. We can conclude that a network fits the power law distribution and displays scale free network properties if $2 < \gamma < 3$, and the p-value of the fitted distribution is $p \geq 0.1$, we also include graphs that only have mean degree, $\bar{d} > 2$ to ensure no sparse graphs are included in the analysis (Broido & Clauset, 2019).

Figure 2.10 summarizes the simulation analysis for all 30 scenarios. Referring back to Table 2.4, scenarios 1-10 represent the variation in node fitness probability, scenarios 11-20 represent the variation in edge fitness probability and 21-30 represent variation in rate of growth of partnerships. (In Appendix 2.2 Figures A2.2 - 12 - 15 summarizes the simulation analysis for the selected scenarios 1, 10, 20 and 30 as representative examples for illustration purposes). The solid colored lines indicate the results obtained from the simulation, we have incorporated a smoothed trend line in black for the variations observed in all 3 graphs. The X axes represent the simulated scenario, the Y axes represent the percentage of networks that displayed scale free property (i.e., $2 < \gamma < 3$, p-value ≥ 0.1, $2 <$ mean degree). The percentage of scale free networks is calculated in the following manner:

1. Run a scenario for $T = 500$. Calculate metrics such as mean degree and parameters of fitted distribution such as γ and p-value for each of the networks at time $t \in T$.
2. Run the same scenario for 25 iterations. Repeat step 1 for each network at time $t \in T$ for each of the runs.
3. Summarize metrics and determine average number of networks that depict scale free behavior for the scenario (average of scale free networks that emerges from 25 runs * 500 time periods per scenario).

Figure 2.10 shows that the variation in node fitness probability results in an increase in the percentage of scale free networks, but diminishes as the range of the distribution increases beyond a certain threshold. Since the node fitness probability of two firms, i and j, have a direct correlation

[8] See Appendix A for detailed scale free graphs of the simulated networks.

on the edge fitness probability between them, e_{ij}, after $t > 1$, nodes with greater fitness probability act as central nodes and create and scaling effect in the network. However, as the range gets wider the central nodes become agglomerates in the network at a faster rate. Hence, the network starts becoming pretty dense more rapidly, not adhering to the scale free properties. That results in a downward trend as the bounds of the node fitness probability distribution increases. This tells us that in an alliance network, having firms with a broader range of characteristics has a direct correlation in resulting in more scale free networks. However, if that range is extremely broad, the alliance creation process gets centered around central players and makes the alliance network extremely dense across firms.

Relative to the increase in the upper bound of the distribution, the initial parameters of the edge fitness probability have relatively no impact on the emergence of scale free property in the network. This is a relatively straightforward implication, since the edge fitness probability distribution directly impacts the propensity to create alliances only at $t = 1$. As explained using Equation 1, after $t > 1$ the edge fitness probability, e_{ij}, is influenced by the node fitness probabilities and the number of alliances that were created between firms i and j by the time $t - 1$. Hence, as seen in Figure 2.9, the initial propensity to create alliances between i and j does not have a long term influence on the emergence of scale free property in these networks. This helps us conclude that among firms, even if there is a large variation of propensity to create alliances at an initial time period $t = 0$, that impact will not hold over time if there are other influences (such as firm-level characteristics and other macro factors) that affect the overall propensity to create alliances.

The variation of partnership growth, has an almost linear upward trending relationship with the number of scale free networks generated as the upper bound of the number of alliances that can be created increases. Beyond a certain range, the percentage of scale free network plateaus indicating that there is minimal impact on the scale free property of the alliance networks. Since the number of nodes in the state space is restricted to $N = 100$, greater rate of alliance formation tends to create a denser network, that after a specific threshold does not adhere to the scale free property. This helps us conclude that, the rate at which alliance formation takes place (due to various macro factors), does have a direct correlation on making an alliance network scale free.

However, as this rate increases while there will still be alliances created at a faster rate within the network, the number of scale free networks that emerge over time will not be impacted.

Hierarchy Property.

As described earlier, in order to evaluate whether near decomposability emerges in the simulated networks, we need to determine if they display hierarchy in networks. In that section, details of how hierarchy emerges when relationship $C(k) \sim k^{-\beta}$ for some stochastic β (Ravasz & Barabási, 2003). In line with this theory, we test the impact of each scenario in the simulated networks to determine to what degree they depict hierarchical behavior.[9]

Figure 2.11 includes the percentage of networks that show hierarchical property on the right and the percentage of networks that display scale free property on the right. As illustrated in the previous section, the existence of both, scale free property and hierarchy, is a necessary condition for near decomposability. We followed the same process detailed out for measuring the percentage of nodes with scale free property for hierarchy as well. Scarce networks were eliminated from the analysis, and we assume there will exist no perfect hierarchical network in the real world. Hence, in $C(k) \sim k^{-\beta}$ for some β, we consider the range for beta to be less than or equal to 2.5 in order to calculate the percentage networks that display hierarchy in the simulated networks.

We observed that the scale free property increases with the increase in range in the node fitness probability, and then decreases as that range increases beyond a certain threshold. However, as seen in Figure 2.10, the increased range in node fitness probability has slowly decreasing effect on the percentage of networks that display hierarchical property. The underlying cause is that beyond a certain point in a network, regardless of the range of the node probability distribution, the network becomes overly clustered and this results in a non-hierarchical network (Appendix 3 - note the loss of linearity from t = 100 to t = 500 in scenarios 1, 5 and 10).

The observation above, in node fitness probability, holds true in the case of varying edge fitness probability as well. The percentage of networks that display hierarchical property on average remains in the range of 10% to 20%, generally as a slow increasing function. Again, when an alliance network's clustering (when firms start creating more and more alliances with the neighbors of firms they already have partnerships with) gets over a certain threshold, we observe

[9] Please refer to Appendix A for graphs on clustering vs. degree for the simulated scenarios).

that the hierarchical property diminishes (Appendix A - note the loss of linearity from t = 100 to t = 500 in scenarios 11, 15 and 20).

In contrast to the variation in node fitness probability and edge fitness probability, the rate of growth of partnerships has a more pronounced impact on the percentage of networks that display hierarchical property. As the upper bound on the distribution for rate of growth of partnerships increases, the percentage of networks that display hierarchical property diminishes significantly. This shows that, the rate of which partnerships grow (and hence create more inter-connections within the alliance network) has a more significant impact on the emergence of hierarchy compared to the firm-level or alliance-level characteristics (Appendix A note the loss of linearity from t = 100 to t = 500 in scenario 21 and 25 and almost non-existence of linear relationship in scenario 30).

2.6.6 Additional Analyses

We conducted multiple additional analyses to evaluate the impact on the simulation results explained above. In this section, we highlight two selected analyses for reference. The first, analyzes the impact of differing node fitness distribution parameters on selected scenarios between scenarios 11-30. This shows the impact of a change in node fitness distribution parameters on the select scenarios compared to the baseline that was analyzed earlier. The second analysis, is a heat map generated to evaluate the stable time periods where the scale free property emerges. This gives a high-level understanding of the time it takes for scale free property to emerge in a simulated network and the length of time at which it remains stable, giving an additional layer of understanding of the pattern in which the property emerges beyond merely the percentage of time periods at which it shows scale freeness.

Sensitivity Analysis under Differing Node Fitness Conditions

In Figure 2.9, we observe that scenario 5 results in the highest percentage of networks with scale free property. Since scenarios 11-30 were tested under conditions where the node fitness ranged between [1,2], additional analysis was conducted to test the variation in results if the node fitness ranged between [1,5], which represents the node fitness distribution in scenario 5. This gives a comparison to baseline where we tested a node distribution of [1,2] against scenarios 11-30. We tested a selected set of scenarios: 13,18, 23 and 28. To recollect, in scenarios 13 and 18 the edge fitness values changes from [1,3] to [1,8]. These two scenarios provide a test case of the impact of increasing the upper bound of the node fitness value from [1,2] to [1,5] on already tested scenarios for varying edge fitness values. In scenarios 23 and 28 the growth probability distribution values change from [1,4] – [1,2] to [1,9] – [1,2]. Thus, these two latter scenarios provide a test case of the impact of increasing the upper bound of the node fitness value from [1,2] to [1,5] on already tested scenarios for varying growth probability distribution values.

In Figure 2.11, the scatter plots represent the percentage of scale free networks when tested under conditions where the node fitness is [1,5] instead of [1,2]. We observe that in scenario 13, the total percentage of scale free networks increases from 24% to 39%, similarly scenario 18 also shows an increase from 22% to 42% when the node fitness distribution increases from [1,2] to [1,5]. This shows that the scale free property is positively impacted under increased node fitness distribution conditions when compared with varying edge fitness parameters.

When tested for scale free property under differing growth distribution scenarios we observe the following. In scenario 23, the total percentage of scale free networks increases significantly from 13% to 46%. But, in scenario 28 we see a decrease from 41% to 27% when the node fitness distribution increases from [1,2] to [1,5]. This shows that the scale free property could be positively or negatively impacted on certain conditions of growth distribution, which will require further detailed analysis to understand if there is a discernible pattern.

The scatter plots in Figure 2.12, represent the percentage of hierarchical networks when tested under the condition where the node fitness is [1,5]. We observe that in scenario 13, the total percentage of hierarchical networks increases from 6% to 15%, whereas scenario 18 shows a decrease from 18% to 4% when the node fitness distribution increases from [1,2] to [1,5]. Unlike the impact of increased node fitness distribution on scale free property, scenarios 13 and 18 do not show a discernible pattern compared to base line for hierarchical property under increased node fitness distribution values. When tested for varying growth distribution scenarios we observe the that in scenario 23, the total percentage of hierarchical networks decreases slightly from 15% to 13%. But, in scenario 28 we see an increase from 5% to 9% when the node fitness distribution increases from [1,2] to [1,5]. Again, for certain conditions of growth distribution the hierarchical property could be positively or negatively impacted, which as mentioned earlier will require further detailed analysis to understand if there is a discernible overall pattern under various growth distribution conditions.

Heat Map to Observe Emergence of Scale Free Property

In Figure 2.13, we display the emergence of scale free property over time. The Y-axis represents the time period in each simulation, ranging from *t=1* to *t=500.* The X-axis represents each of the scenarios. Each scenario was run for 25 iterations and the color coding indicates the percentage of iterations in which the scale free property emerged for each scenario.

We observe that under varying node fitness conditions (scenarios 1-10), the scale free property emerges only after scenario 3, indicating a minimum threshold that is required for node distribution values so that the scale free property emerges. After scenario 3, when the range in the node distribution values become greater, scale free property begins to emerge earlier in the time horizon and remains stable for a set time range. As the node fitness distribution range increases, the stability period, indicating how long the scale free property remains stable, decreases. In scenario 3 we observe the stable period to range from approximately between *t=150* to *t=500* whereas, by scenario 10, the stable period shrinks from approximately *t=50* to *t=200.*

For scenarios 11-20, where the edge fitness varies, the scale free property emerges approximately after *t=200* in some scenarios, and after *t=300* in most scenarios. Thereafter, the stable condition remains up until *t=500,* of these 10 scenarios for 5 of them over 50% of the iterations had scale free property emerge after approximately *t=450*.

For scenarios 21-30, where the growth distribution varies, the scale free property does not emerge until the growth distribution reaches a threshold of [1,4]-[1,2] in scenario 23. From scenario 23 onwards, scaling property emerges after approximately *t=100,* and remains stable until the end of the time horizon. In general, the greater the growth distribution, the greater the number of iterations that displayed scaling property across many time periods.

In summary, this analysis shows that the emergence of scaling property across the time horizon varies depending on the scenario. When node fitness is varied, the length of the stable condition peaks between scenarios 3 and 6, while when the edge fitness and growth distribution is varied, the stable condition remains approximately the same after a certain threshold.

2.7 Conclusion and Future Research

2.7.1 Summary of the Research

In section 2, we hypothesized in interfirm networks, smaller network sub communities recursively organized hierarchically into a larger network in a nested fashion. These nested subcommunities emerging from the hierarchical network exhibit local-level clustering. Sections 3 and 4, in order, explored empirical evidence in inter-firm alliance networks, and simulated the generation of alliance networks under various conditions to test the above hypothesis.

From an empirical standpoint, we observe that inter-firm alliances, as observed in the expansive biopharmaceutical alliance data, develop near decomposable network characteristics over time. Considering a rolling 5 and 3 year window, we show that biopharmaceutical alliances develop scale free property and hierarchical property over time, which makes them near decomposable systems. Using a simulation, we attempt to emulate firm-level, dyadic relationship level and macro level scenarios that will help us understand what underlying characteristics impact the emergence of near decomposability. We observe that firm-level characteristics in the form of node fitness probability does have an impact on the emergence of scale free property in alliance networks, but its variation has no consequential impact on the emergence of networks with hierarchical property. Hence, the overall percentage of networks in a given time horizon that show near decomposability remain on average between 10% - 20% in our simulated alliance environment. However, the varying fitness in the dyadic relationship in the form of edge fitness probability does not have a significant impact on scale free property and hierarchical structure. Again, in this case the overall percentage of near decomposable networks remain at an average between 10% and 20%. Finally, the network-wide propensity to create alliances, rate of growth of partnerships, does positively influence the emergence of scale free networks, but negatively influences the emergence of hierarchical network structures. This research has paid attention to not only develop an empirical understanding of near decomposability in inter-firm networks but also simulated the firm-specific, alliance specific and macro level heterogeneity observed in the real world.

2.7.2 Research Contributions

This paper makes a number of contributions to the literature. First, we develop both conceptually and empirically an approach to think about near decomposability in interorganizational settings. In doing so, we offer a pathway to think about the implications of business partnerships both at the industry and firm levels. If, as we show, that near decomposability can emerge under specific circumstances and there are implications at the industry-level, both the business and government policies would want to pay attention to possibly enabling those circumstances. For example, as during the case of COVID-19, governments mandating partnerships for the broader welfare circumstances (E.g., Johnson & Johnson partnership with Merck to manufacture vaccine), industrial policies may enable structural circumstances conducive for the emergence and growth of the industries. We hope that the theoretical foundation developed in the paper will provide opportunities on how near decomposability can be helpful for wide variety of performance and evolutionary outcome measures.

Similarly, an expansive literature has long recognized the firm-specific implications of alliances and networks, and how they affect the process of these firms create and capture value (Ahuja *et al.*, 2012; Gulati, 1998; Gulati *et al.*, 2000; Mesquita *et al.*, 2017). How firms embed themselves in a network "[..] is a logic of exchange that promotes economies of time, integrative agreements, Pareto improvements in allocative efficiency, and complex adaptation" (Uzzi, 1997). While articles in this nature provide evidence for complex interactions among firms, broader systems-wide mechanisms with nonsimple interactions remains broadly an open question and this paper attempts to provide a first step in answering the question. By showing how firms' distinct heterogeneous characteristics can emerge from the overall nature of near decomposability in the network, we add to the literature that identifies the sources competitive advantage firms develop, sustain and grow during the evolutionary processes (Baum *et al.*, 2003; Borgatti & Foster, 2003; Rowley *et al.*, 2000; Schilling & Phelps, 2007; Uzzi & Spiro, 2005). This also helps us to more broadly contribute to the growing stream of research that takes a structural approach to network studies. This strand of research elucidates that substantial causal drivers in network settings often come from the enduring patterns of relationships among the actors in the network (Kilduff & Brass, 2010; Wellman & Berkowitz, 1988). By providing a treatment on how a macro-level hierarchy we observe in the network may emerge from micro-level local clustering and the scaling in network, we contribute to this stream of literature.

An important challenge in identifying the patterns of clustering in strategic networks and their network evolutionary behavior is empirical. Ahuja *et al.* (2012) articulates that the "paucity of empirical research likely stems from challenges such as the practical difficulties posed by obtaining longitudinal network data, the complexities of handling networks over time [..] (f)or the field to advance, a cumulative body of empirical evidence is needed to advance our understanding about the emergence, evolution, and dynamics of networks". The most obvious methodological challenge here arises from the fact that the typical toolkits of organizational and strategy scholars are not the best to identify network patterns. State-of-the-art network science techniques, that has made significant inroad in the recent years, typically emerge from scientific disciplines such as computer science, engineering and mathematics (Alderson, 2008; Barabási, 2016; Newman & Girvan, 2004). Naturally these potential methods for solving these puzzles are intellectually distant from the status quo organizational scholarship. These methodological advancements help us to capture some the real world network attributes and their potential implications with a reasonable degree of confidence.

2.7.3 Limitations and Future Research

Several limitations in this study point to future research directions. First, the simulation model developed in this research, has been tested purely for node fitness, edge fitness and rate of growth of partnerships. These factors are generalized to represent firm-level, dyadic and macro market influencing factors that drive alliance creation. An implicit assumption in the paper is that industry dynamics is cohesive and evolutionary process could be studied within a specific industry. While reasonably cohesive nature of biopharmaceutical industry allows us to make this assumption and it gives a valuable high level insight into the factors that impact the emergence of near decomposability in alliance partnerships, future work focus on modeling the evolutionary process more exhaustively to capture historical trends in alliance partnerships of firms operating in multiple industries. This would require a study of detailed alliance data spanning across multiple industries, which is much the scope of this paper.

Second, we are limit our simulation analysis only to rate of growth when thinking about the environmental factors. Business partnerships are both facilitated and hampered by several environmental factors including the institutional, cultural, legal and other social environments. Although limiting the scope of our study helps to tease out some specific mechanisms of interest for this paper, future work should consider relaxing this assumption and explicitly model the contingencies it may bring to the formation of near decomposable property in networks

Third, we test a generalized network propagation mechanism using where the sample is drawn from a probability-weighted uniform distribution in Monte Carlo simulation. However, testing more empirically driven propagation mechanisms might give insights into their impact on the emergence of near decomposable networks. Defining real world alliance network characteristic based on historical observations of firm-level and dyadic-level characteristics will be an interesting expansion of this simulation.

Fourth, in this paper we primarily focus on network propagation and formation of alliances. However, an important next step is to study how the existence of near decomposability, or lack thereof, will affect both traditional firm- and alliance-level performance measures such as financial outcomes, R&D performance and innovation measures. We leave these ideas for future research opportunities.

2.8 References

Adegbesan JA, Higgins MJ. 2011. The Intra-Alliance Division of Value Created through Collaboration. *Strategic Management Journal* **32**(2): 187-211.

Ahuja G. 2000. Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study. *Administrative science quarterly* **45**(3): 425-455.

Ahuja G, Soda G, Zaheer A. 2012. The Genesis and Dynamics of Organizational Networks. *Organization science* **23**(2): 434-448.

Albert R, Barabási A-L. 2002. Statistical Mechanics of Complex Networks. *Reviews of modern physics* **74**(1): 47.

Alderson DL. 2008. Or Forum—Catching the "Network Science" Bug: Insight and Opportunity for the Operations Researcher. *Operations research* **56**(5): 1047-1065.

Barabási A-L. 2016. *Network Science*. Cambridge University Press.

Barabasi A-L, Oltvai ZN. 2004. Network Biology: Understanding the Cell's Functional Organization. *Nature reviews genetics* **5**(2): 101-113.

Barabási A-L, Ravasz E, Oltvai Z. 2003. Hierarchical Organization of Modularity in Complex Networks. In *Statistical Mechanics of Complex Networks.* Springer.

Barney JB. 1996. The Resource-Based Theory of the Firm. *Organization science* **7**(5): 469-469.

Baum JA, Calabrese T, Silverman BS. 2000. Don't Go It Alone: Alliance Network Composition and Startups' Performance in Canadian Biotechnology. *Strategic management journal* **21**(3): 267-294.

Baum JA, Shipilov AV, Rowley TJ. 2003. Where Do Small Worlds Come From? *Industrial and Corporate change* **12**(4): 697-725.

Baumann O. 2015. Distributed Problem Solving in Modular Systems: The Benefit of Temporary Coordination Neglect. *Systems Research and Behavioral Science* **32**(1): 124-136.

Berg S, Friedman P. 1978. Technological Complementarities and Industrial Patterns of Joint Venture Activity, 1964-1975. *Industrial Organization Review* **6**: 110-116.

Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U. 2006. Complex Networks: Structure and Dynamics. *Physics reports* **424**(4-5): 175-308.

Borgatti SP, Foster PC. 2003. The Network Paradigm in Organizational Research: A Review and Typology. *Journal of management* **29**(6): 991-1013.

Borgatti SP, Halgin DS. 2011. On Network Theory. *Organization science* **22**(5): 1168-1181.

Broido AD, Clauset A. 2019. Scale-Free Networks Are Rare. *Nature communications* **10**(1): 1-10.

Burt RS. 2001. Closure as Social Capital. *Social capital: Theory and research*: 31-55.

Burt RS. 2005. *Brokerage and Closure: An Introduction to Social Capital*. Oxford university press.

Burt RS. 2009. *Structural Holes: The Social Structure of Competition*. Harvard university press.

Clauset A, Newman ME, Moore C. 2004. Finding Community Structure in Very Large Networks. *Physical review E* **70**(6): 066111.

Coleman JS. 1988. Social Capital in the Creation of Human Capital. *American Journal of Sociology* **94**(s).

Coleman JS. 1994. *Foundations of Social Theory*. Harvard university press.

Dhanaraj C, Parkhe A. 2006. Orchestrating Innovation Networks. *Academy of management review* **31**(3): 659-669.

Dyer JH, Singh H. 1998. The Relational View: Cooperative Strategy and Sources of Interorganizational Competitive Advantage. *Academy of management review* **23**(4): 660-679.

Egidi M, Marengo L. 2004. Near-Decomposability, Organization, and Evolution: Some Notes on Herbert Simon’s Contribution. *Models of a man: Essays in memory of Herbert A. Simon*: 335-350.

Ethiraj SK, Levinthal D. 2004. Modularity and Innovation in Complex Systems. *Management science* **50**(2): 159-173.

Frenken K, Marengo L, Valente M. 1999. Interdependencies, Nearly-Decomposability and Adaptation. In *Computational Techniques for Modelling Learning in Economics.* Springer.

Gavetti G, Levinthal DA, Rivkin JW. 2005. Strategy Making in Novel and Complex Worlds: The Power of Analogy. *Strategic Management Journal* **26**(8): 691-712.

Gilsing V, Nooteboom B, Vanhaverbeke W, Duysters G, Van Den Oord A. 2008. Network Embeddedness and the Exploration of Novel Technologies: Technological Distance, Betweenness Centrality and Density. *Research policy* **37**(10): 1717-1731.

Girvan M, Newman ME. 2002. Community Structure in Social and Biological Networks. *Proceedings of the national academy of sciences* **99**(12): 7821-7826.

Gomes-Casseres B. 1994. Group Versus Group: How Alliance Networks Compete. *Harvard business review* **72**(4): 62-66.

Granovetter M. 1985. Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology* **91**(3): 481-510.

Gulati R. 1998. Alliances and Networks. *Strategic management journal* **19**(4): 293-317.

Gulati R. 1999. Network Location and Learning: The Influence of Network Resources and Firm Capabilities on Alliance Formation. *Strategic management journal* **20**(5): 397-420.

Gulati R, Nohria N, Zaheer A. 2000. Strategic Networks. *Strategic Management Journal* **21**(3): 203-215.

Gulati R, Wohlgezogen F, Zhelyazkov P. 2012. The Two Facets of Collaboration: Cooperation and Coordination in Strategic Alliances. *Academy of Management Annals* **6**(1): 531-583.

Hagedoorn J. 2002. Inter-Firm R&D Partnerships: An Overview of Major Trends and Patterns since 1960. *Research policy* **31**(4): 477-492.

Hamel G. 1991. Competition for Competence and Interpartner Learning within International Strategic Alliances. *Strategic Management Journal* **12**(S1): 83-103.

Harrigan KR, Newman WH. 1990. Bases of Interorganization Co-Operation: Propensity, Power, Persistence. *Journal of Management Studies* **27**(4): 417-434.

Hess AM, Rothaermel FT. 2011. When Are Assets Complementary? Star Scientists, Strategic Alliances, and Innovation in the Pharmaceutical Industry. *Strategic Management Journal* **32**(8): 895-909.

Hite JM, Hesterly WS. 2001. The Evolution of Firm Networks: From Emergence to Early Growth of the Firm. *Strategic management journal* **22**(3): 275-286.

Hitt MA, Ahlstrom D, Dacin MT, Levitas E, Svobodina L. 2004. The Institutional Effects on Strategic Alliance Partner Selection in Transition Economies: China Vs. Russia. *Organization science* **15**(2): 173-185.

Kilduff M, Brass DJ. 2010. Organizational Social Network Research: Core Ideas and Key Debates. *Academy of management annals* **4**(1): 317-357.

Koza MP, Lewin AY. 1998. The Co-Evolution of Strategic Alliances. *Organization science* **9**(3): 255-264.

Laursen K, Reichstein T, Trombini G. Year. Whom to Choose as License Partner? The Interplay between Technology Cooperation & Market Competition. In Proceedings of the Academy of Management Proceedings.

Levinthal DA, March JG. 1993. The Myopia of Learning. *Strategic management journal* **14**(S2): 95-112.

Levinthal DA, Workiewicz M. 2018. When Two Bosses Are Better Than One: Nearly Decomposable Systems and Organizational Adaptation. *Organization Science* **29**(2): 207-224.

Madhavan R, Koka BR, Prescott JE. 1998. Networks in Transition: How Industry Events (Re) Shape Interfirm Relationships. *Strategic management journal* **19**(5): 439-459.

Makadok R. 2003. Doing the Right Thing and Knowing the Right Thing to Do: Why the Whole Is Greater Than the Sum of the Parts. *Strategic Management Journal* **24**(10): 1043-1055.

Mesquita LF, Ragozzino R, Reuer JJ. 2017. *Collaborative Strategy: Critical Issues for Alliances and Networks*. Edward Elgar Publishing.

Mindruta D, Moeen M, Agarwal R. 2016. A Two-Sided Matching Approach for Partner Selection and Assessing Complementarities in Partners' Attributes in Inter-Firm Alliances. *Strategic Management Journal* **37**(1): 206-231.

Newman ME, Girvan M. 2004. Finding and Evaluating Community Structure in Networks. *Physical review E* **69**(2): 026113.

Nickerson JA, Zenger TR. 2004. A Knowledge-Based Theory of the Firm—the Problem-Solving Perspective. *Organization science* **15**(6): 617-632.

Oxley JE. 1999. Institutional Environment and the Mechanisms of Governance: The Impact of Intellectual Property Protection on the Structure of Inter-Firm Alliances. *Journal of Economic Behavior & Organization* **38**(3): 283-309.

Ozmel U, Guler I. 2015. Small Fish, Big Fish: The Performance Effects of the Relative Standing in Partners' Affiliate Portfolios. *Strategic Management Journal* **36**(13): 2039-2057.

Park SH, Chen R, Gallagher S. 2002. Firm Resources as Moderators of the Relationship between Market Growth and Strategic Alliances in Semiconductor Start-Ups. *Academy of management Journal* **45**(3): 527-545.

Powell WW, Koput KW, Smith-Doerr L. 1996. Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology. *Administrative science quarterly*: 116-145.

Powell WW, White DR, Koput KW, Owen-Smith J. 2005. Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. *American journal of sociology* **110**(4): 1132-1205.

Ravasz E, Barabási A-L. 2003. Hierarchical Organization in Complex Networks. *Physical review E* **67**(2): 026112.

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási A-L. 2002. Hierarchical Organization of Modularity in Metabolic Networks. *science* **297**(5586): 1551-1555.

Reinstaller A. 2007. Koen Frenken: Innovation, Evolution and Complexity Theory, Springer.

Reuer JJ, Ragozzino R. 2006. Agency Hazards and Alliance Portfolios. *Strategic management journal* **27**(1): 27-43.

Rivkin JW, Siggelkow N. 2003. Balancing Search and Stability: Interdependencies among Elements of Organizational Design. *Management science* **49**(3): 290-311.

Rowley T, Behrens D, Krackhardt D. 2000. Redundant Governance Structures: An Analysis of Structural and Relational Embeddedness in the Steel and Semiconductor Industries. *Strategic management journal* **21**(3): 369-386.

Rumelt RP, Schendel D, Teece DJ. 1994. *Fundamental Issues in Strategy: A Research Agenda*. Harvard Business School Press.

Salancik GR. 1995. Wanted: A Good Network Theory of Organization, JSTOR.

Schilling MA. 2009. Understanding the Alliance Data. *Strategic Management Journal* **30**(3): 233-260.

Schilling MA, Phelps CC. 2007. Interfirm Collaboration Networks: The Impact of Large-Scale Network Structure on Firm Innovation. *Management science* **53**(7): 1113-1126.

Shan W, Walker G, Kogut B. 1994. Interfirm Cooperation and Startup Innovation in the Biotechnology Industry. *Strategic management journal* **15**(5): 387-394.

Simon H. 1962. A. 1962. The Architecture of Complexity. *Proceedings of the American Philosophical Society* **106**(6): 467-482.

Simon HA. 1977. The Organization of Complex Systems. In *Models of Discovery.* Springer.

Simon HA. 1996. The Sciences of the Artificial (Vol. 136), MIT press.

Simon HA. 2000. Can There Be a Science of Complex Systems. *Unifying themes in complex systems* **1**: 3-14.

Simon HA. 2002. Near Decomposability and the Speed of Evolution. *Industrial and corporate change* **11**(3): 587-599.

Stern I, Dukerich JM, Zajac E. 2014. Unmixed Signals: How Reputation and Status Affect Alliance Formation. *Strategic Management Journal* **35**(4): 512-531.

Stuart TE. 1998. Network Positions and Propensities to Collaborate: An Investigation of Strategic Alliance Formation in a High-Technology Industry. *Administrative science quarterly*: 668-698.

Uzzi B. 1996. The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect. *American sociological review*: 674-698.

Uzzi B. 1997. Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness. *Administrative science quarterly*: 35-67.

Uzzi B, Spiro J. 2005. Collaboration and Creativity: The Small World Problem. *American journal of sociology* **111**(2): 447-504.

Walker G, Kogut B, Shan W. 1997. Social Capital, Structural Holes and the Formation of an Industry Network. *Organization science* **8**(2): 109-125.

Wellman B, Berkowitz SD. 1988. *Social Structures: A Network Approach*. CUP Archive.

Wu J. 2013. Hierarchy Theory: An Overview. *Linking ecology and ethics for a changing world*: 281-301.

Yang H, Zheng Y, Zhao X. 2014. Exploration or Exploitation? Small Firms' Alliance Strategies with Large Firms. *Strategic Management Journal* **35**(1): 146-157.

Yayavaram S, Ahuja G. 2008. Decomposability in Knowledge Structures and Its Impact on the Usefulness of Inventions and Knowledge-Base Malleability. *Administrative Science Quarterly* **53**(2): 333-362.

Zaheer A, Bell GG. 2005. Benefiting from Network Position: Firm Capabilities, Structural Holes, and Performance. *Strategic management journal* **26**(9): 809-825.

Zaheer A, Soda G. 2009. Network Evolution: The Origins of Structural Holes. *Administrative Science Quarterly* **54**(1): 1-31.

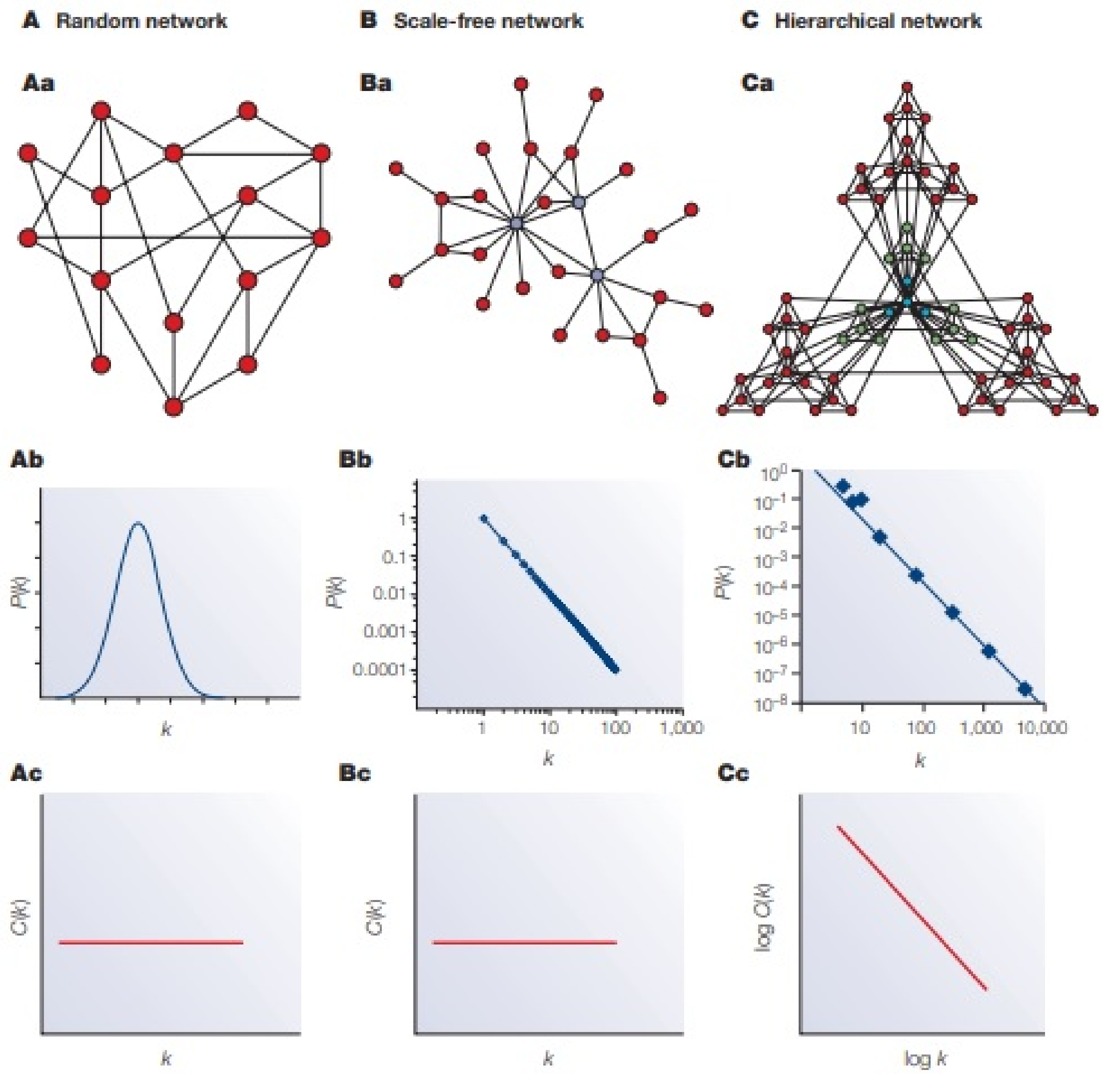


Figure 2.1 Comparison of Random, Scale-Free and Hierarchical Network

(Figure is reproduced from Barabasi and Oltvai (2004) with permission.)

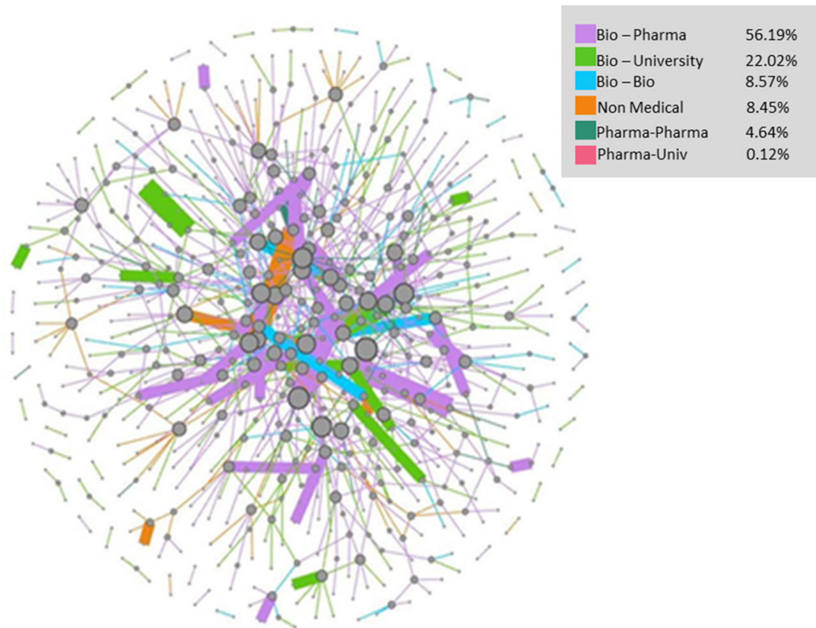


Figure 2.2 Network Descriptive Plots – Years 1985 – 1989

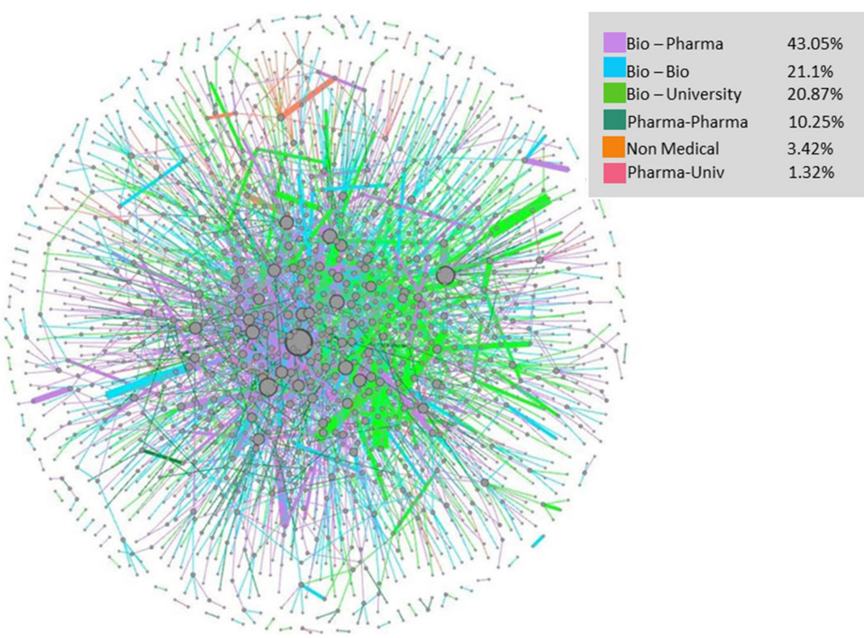


Figure 2.3 Network Descriptive Plots – Years 1991 – 1995

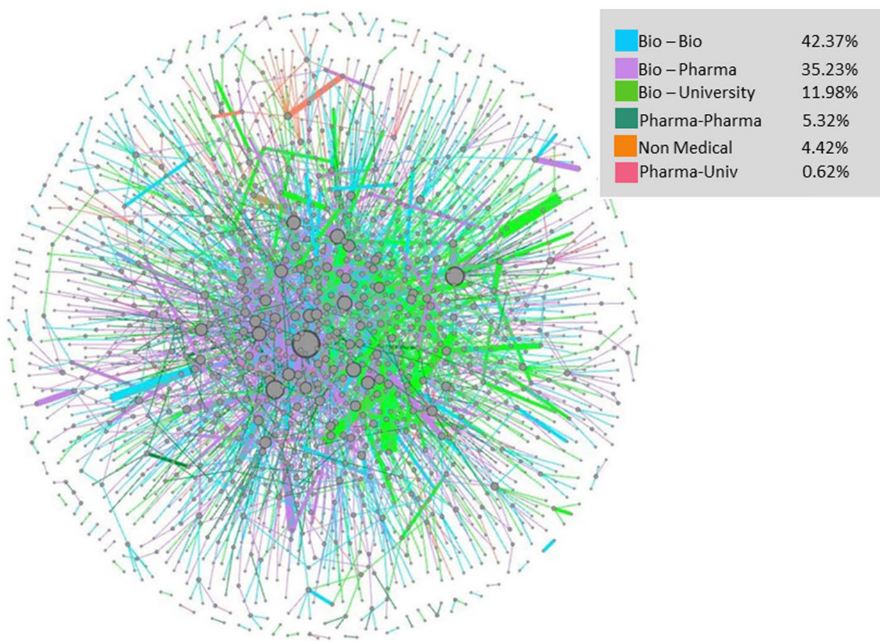


Figure 2.4 Network Descriptive Plots – Years 1997 - 2001

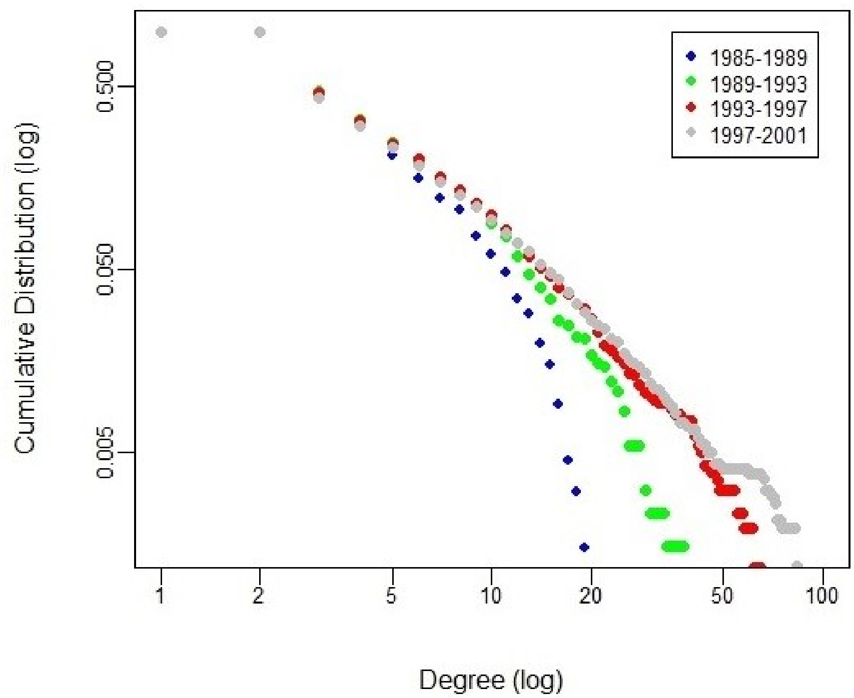


Figure 2.5 Degree Distribution of the Network, 5 Year Period (in log – log)

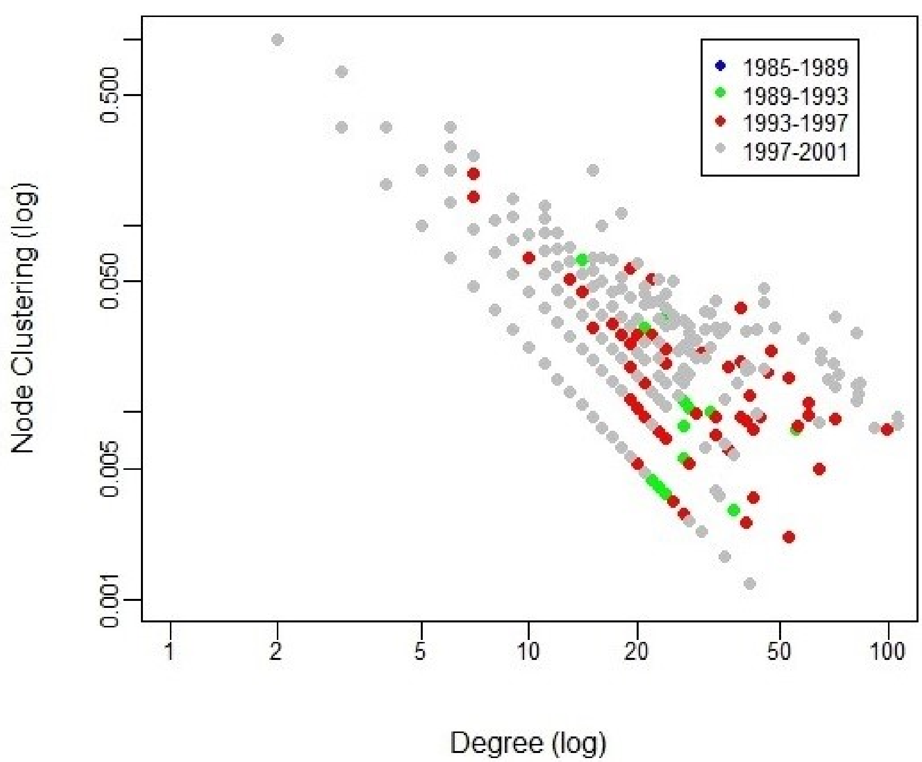


Figure 2.6 Hierarchical Structure of the Network, 5 Year Period (in log – log)

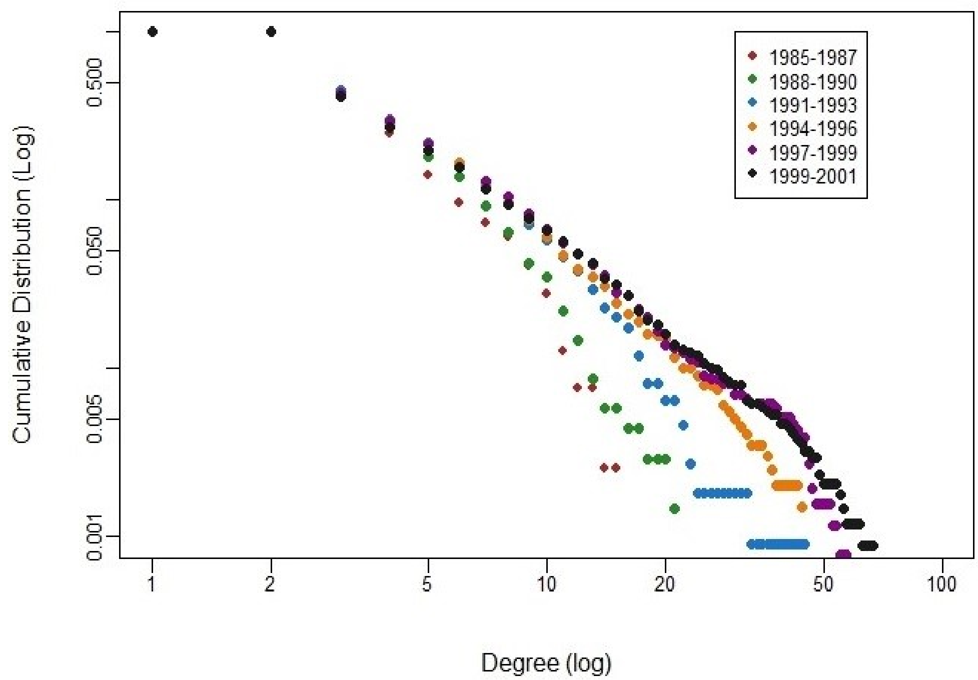


Figure 2.7 Hierarchical Structure of the Network, 3 Year Period (in log – log)

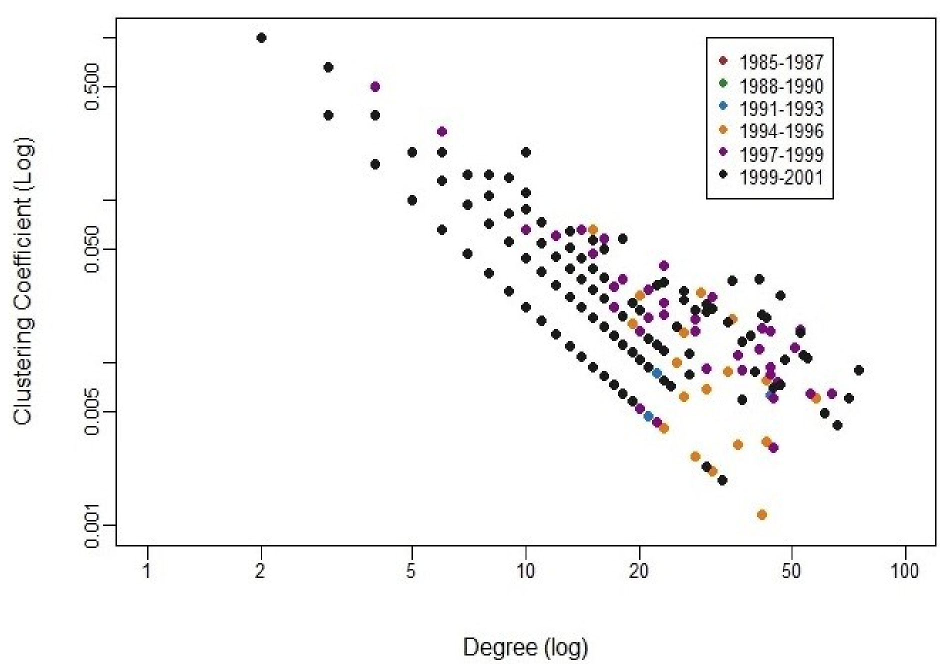


Figure 2.8 Hierarchical Structure of the Network, 3 Year Period

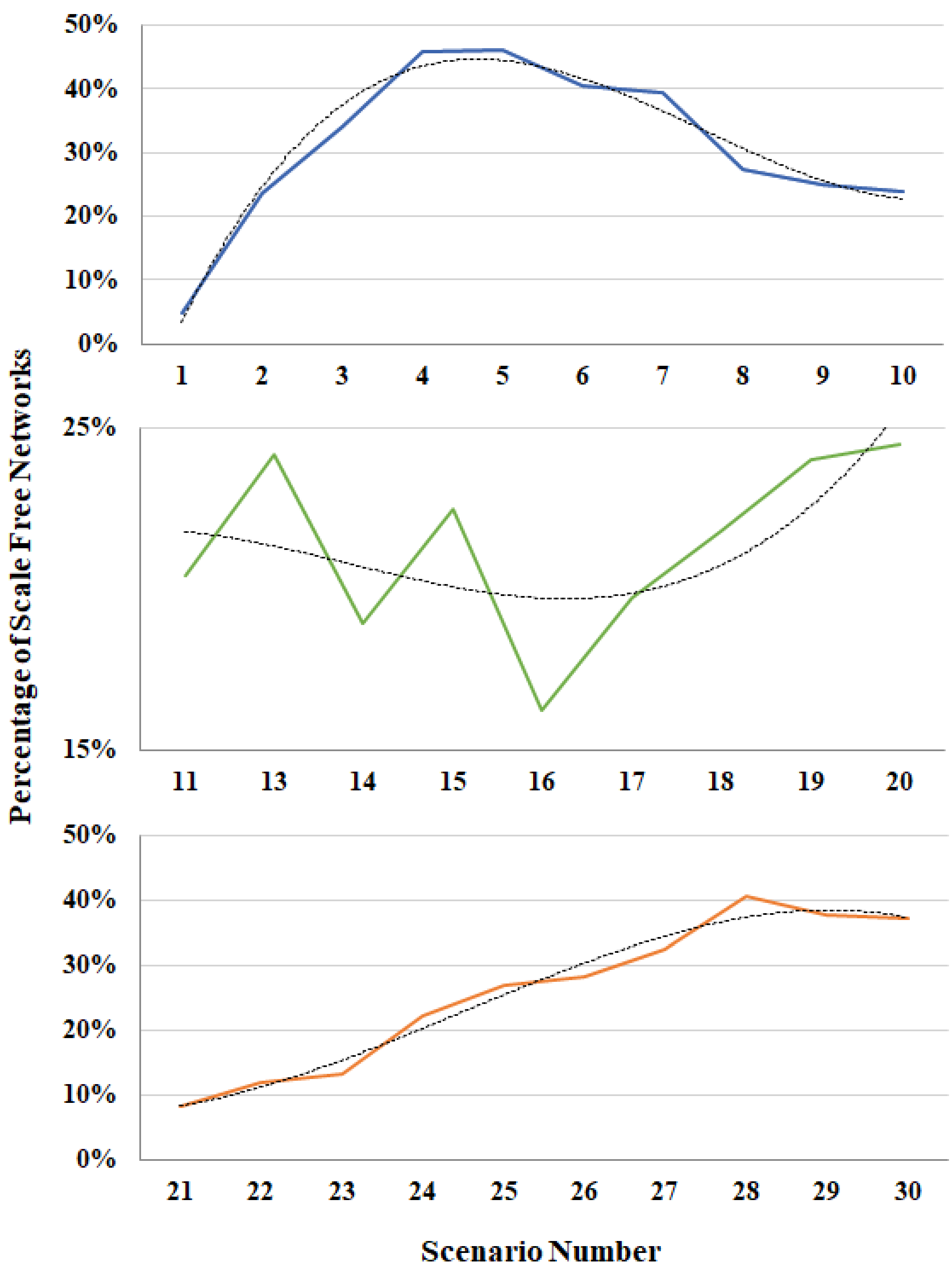


Figure 2.9 Percentage of Networks that Display Scale Free Property by Scenario

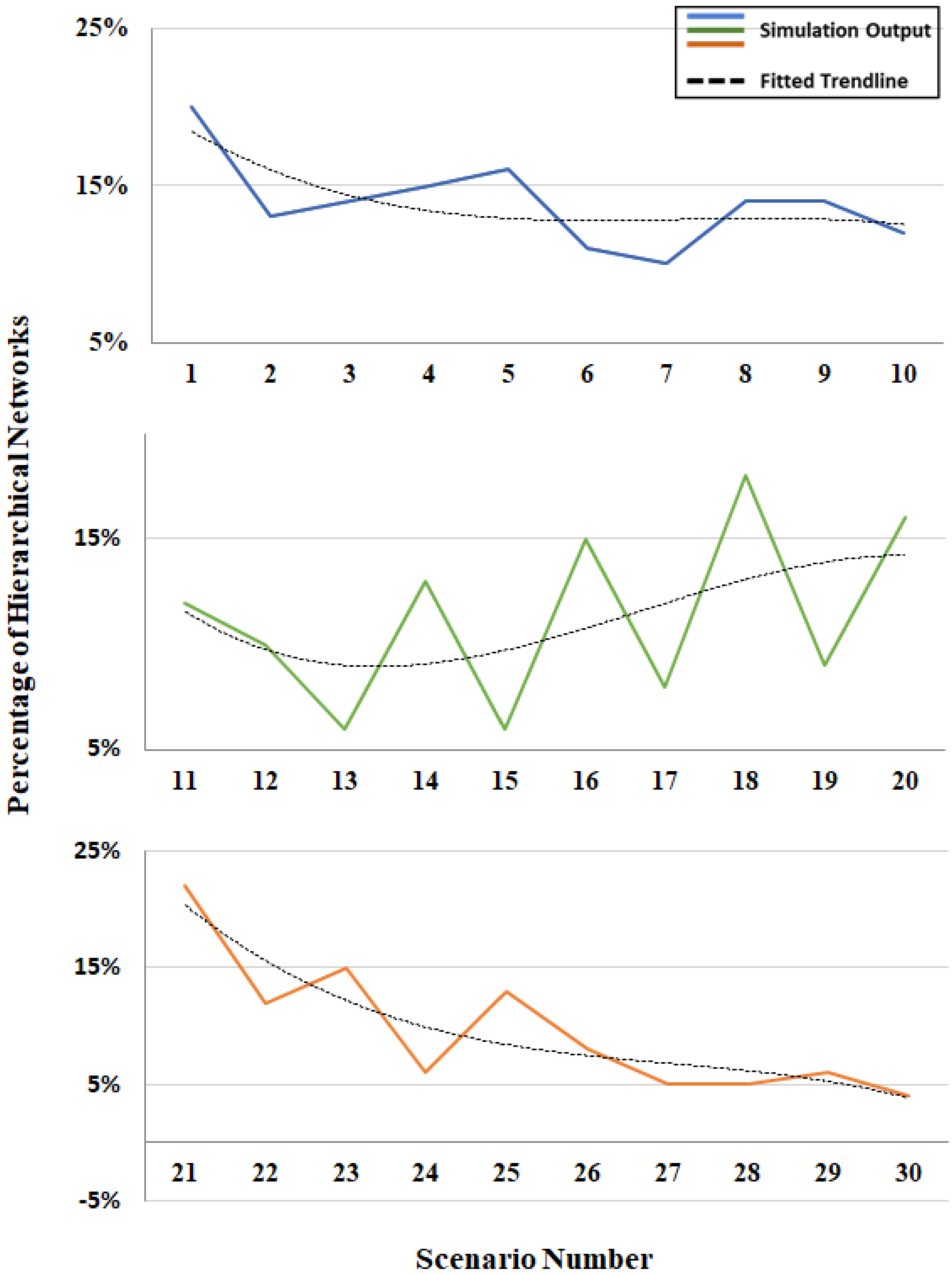


Figure 2.10 Percentage of Networks with Hierarchy Property by Scenario

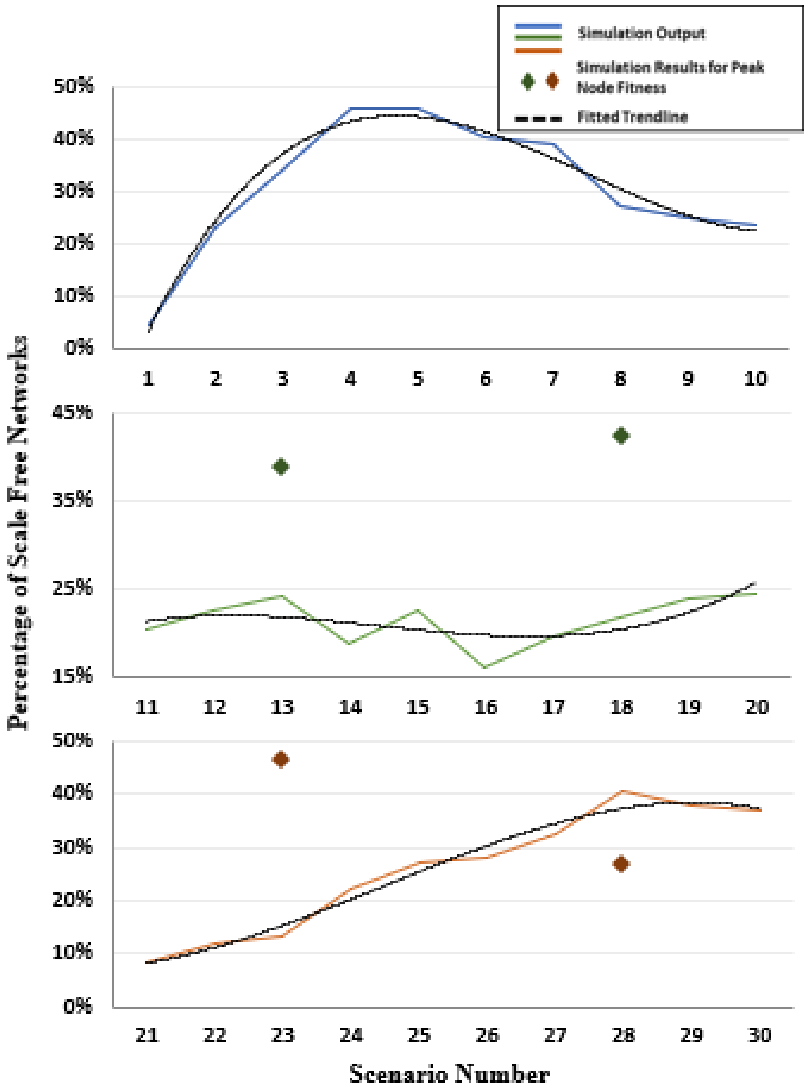


Figure 2.11 Percentage of Networks with Scale Free Property Under Differing Node Fitness Conditions for Scenarios 13,18, 23 and 28

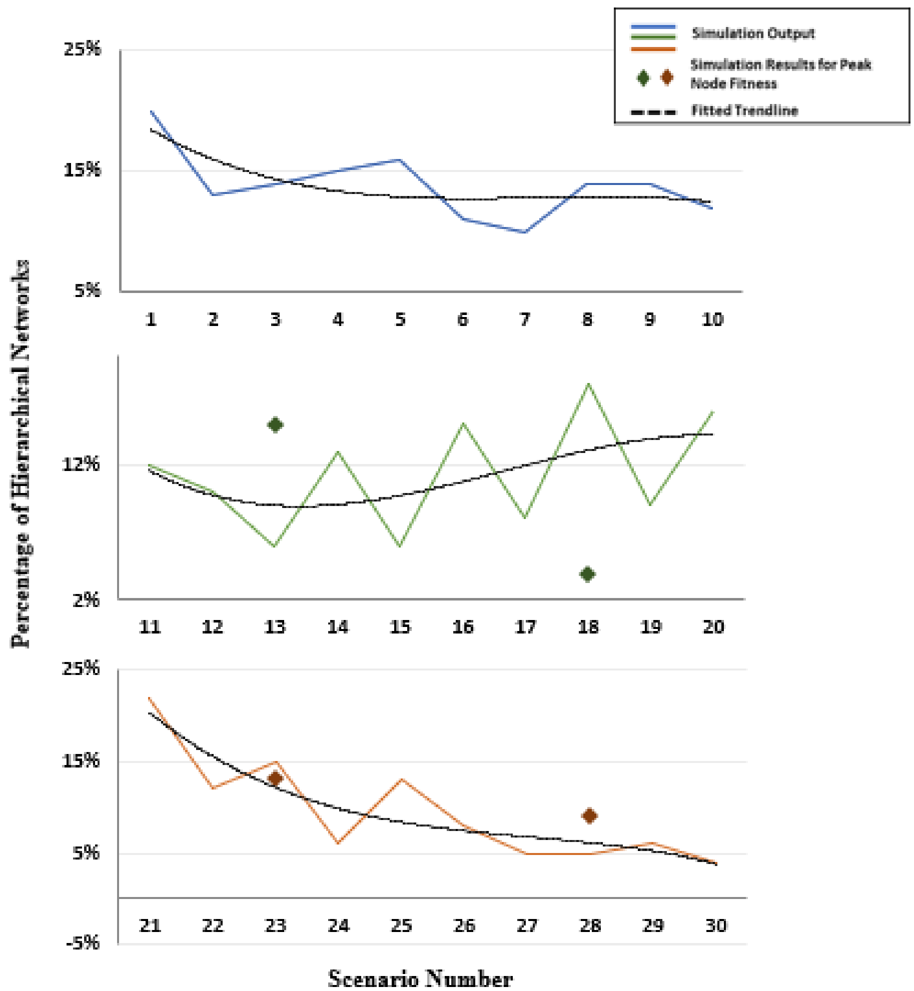


Figure 2.12 Percentage of Networks with Hierarchical Property Under Differing Node Fitness Conditions for Scenarios 13,18, 23 and 28

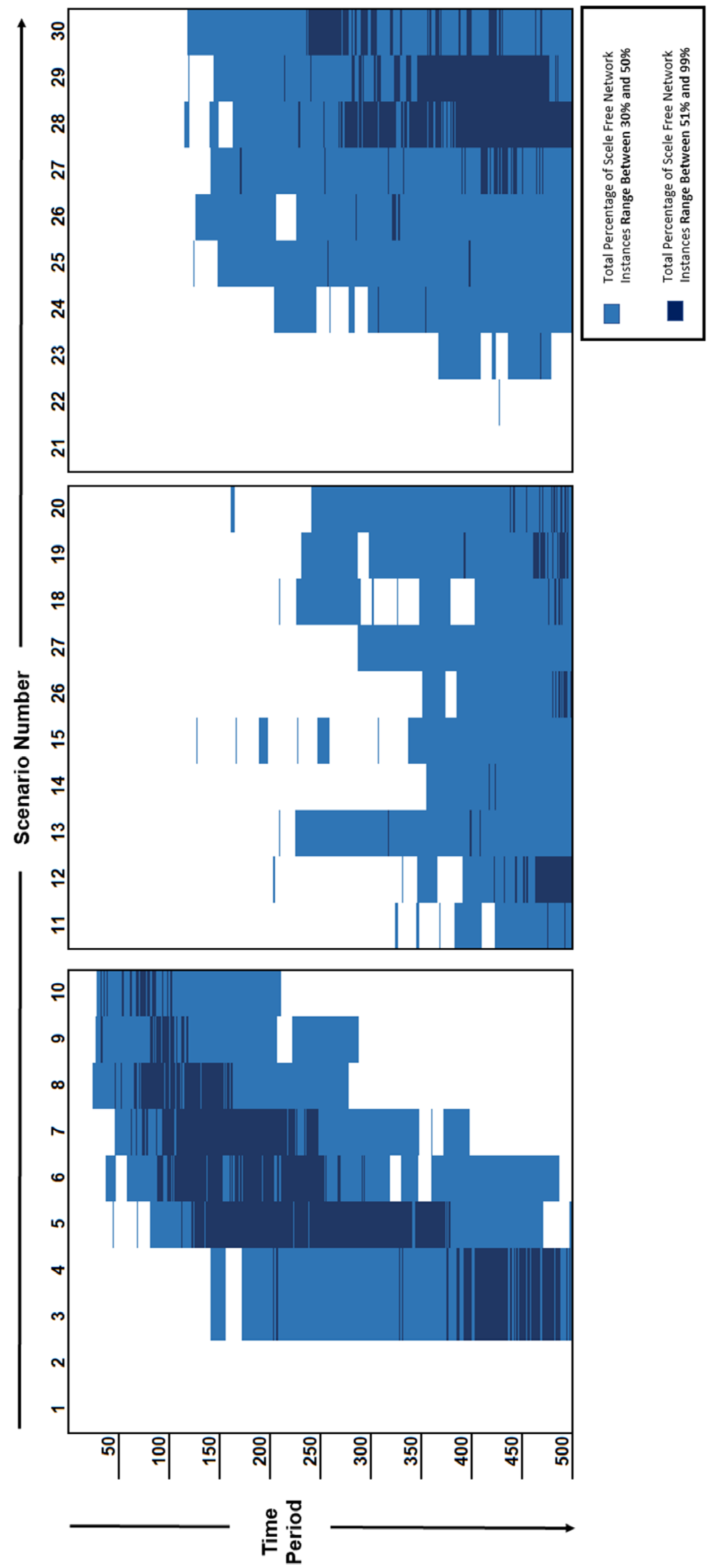


Figure 2.13 Heat Map of Emergence of Scale Free Property

Table 2.1 Alliance Types – Changes over Time

		Bio-Pharma	**Bio-University**	**Bio-Bio**	**Non-Medical**	**Pharma-Pharma**	**Pharma-University**
Percentage of Overall Alliance During Period (%)	1985-1989	56.19	22.02	8.57	8.45	4.64	0.12
	1986-1990	55.21	22.25	10.51	7.55	4.30	0.19
	1987-1991	54.26	23.97	11.91	6.39	3.31	0.16
	1988-1992	48.60	25.22	16.03	5.62	4.34	0.19
	1989-1993	45.26	24.24	17.92	4.82	7.51	0.26
	1990-1994	43.22	23.03	19.37	3.95	9.79	0.66
	1991-1995	43.05	20.87	21.10	3.42	10.25	1.32
	1992-1996	42.13	18.48	23.82	3.35	10.78	1.44
	1993-1997	42.94	16.31	25.93	3.29	10.14	1.40
	1994-1998	42.44	14.81	29.48	3.51	8.44	1.32
	1995-1999	42.05	12.96	33.20	3.64	7.05	1.10
	1996-2000	38.21	12.59	38.37	4.19	5.90	0.74
	1997-2001	35.23	11.98	42.37	4.42	5.32	0.68

Table 2.2 KS-Statistic, 5 Year Period

5 Year Periods	Beta	KS Statistic	P-Value
1985-1989	10.03	0.07	1.00
1986-1990	7.45	0.06	1.00
1987-1991	7.78	0.08	1.00
1988-1992	3.37	0.06	0.88
1989-1993	3.48	0.05	0.93
1990-1994	3.02	0.06	0.69
1991-1995	3.39	0.06	0.94
1992-1996	3.04	0.03	1.00
1993-1997	2.83	0.04	0.83
1994-1998	2.88	0.03	0.98
1995-1999	3.06	0.04	0.99
1996-2000	2.99	0.04	0.98
1997-2001	2.98	0.04	0.98

Table 2.3 KS-Statistic, 3 Year Period

3 Year Periods	Beta	KS Statistic	P-Value
1985-1987	2.87	0.06	0.83
1986-1988	8.76	0.10	1.00
1987-1989	8.19	0.04	1.00
1988-1990	5.20	0.04	1.00
1989-1991	4.26	0.06	0.97
1990-1992	5.41	0.04	1.00
1991-1993	3.04	0.06	0.66
1992-1994	3.11	0.06	0.86
1993-1995	2.94	0.04	0.97
1994-1996	2.91	0.04	0.97
1995-1997	3.21	0.05	0.97
1996-1998	2.69	0.04	0.77
1997-1999	2.77	0.04	0.83
1998-2000	3.14	0.04	0.99
1999-2001	2.90	0.04	0.92

Table 2.4 Definition of Simulation Parameters

Notation	Definition
N	Total number of nodes (firms) in the state space of the alliance network
T	Total number of periods in the time horizon
e_{ij}	Edge probability (alliance probability) between nodes (firms) $i, j \in N$
β_i	Node fitness probability of node (firm) $i \in N$
α_{ij}	Number of edges (alliances) between nodes (firms) $i, j \in N$
A	Edge (alliance) fitness matrix (A is an N*N Matrix)
p_t	Number of edges (alliances) to be created in time period $t \in T$
q_t	Number of edges (alliances) to be discontinued in time period $t \in T$

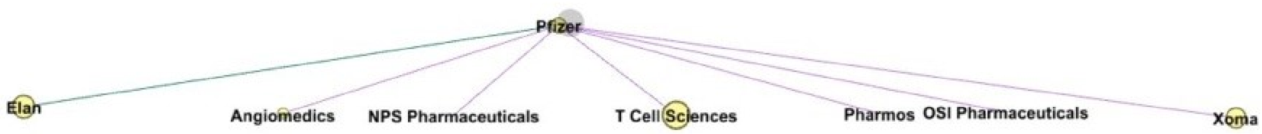
Table 2.5 Summary of Scenarios Tested in the Simulation

Scenario	Node Fitness Prob Dist.	Edge Fitness Prob Dist.	Growth Prob Dist.
Scenario 1	**[1,1]**	[1,2]	[1,5] - [1,2]
Scenario 2	**[1,2]**	[1,2]	[1,5] - [1,2]
Scenario 3	**[1,3]**	[1,2]	[1,5] - [1,2]
Scenario 4	**[1,4]**	[1,2]	[1,5] - [1,2]
Scenario 5	**[1,5]**	[1,2]	[1,5] - [1,2]
Scenario 6	**[1,6]**	[1,2]	[1,5] - [1,2]
Scenario 7	**[1,7]**	[1,2]	[1,5] - [1,2]
Scenario 8	**[1,8]**	[1,2]	[1,5] - [1,2]
Scenario 9	**[1,9]**	[1,2]	[1,5] - [1,2]
Scenario 10	**[1,10]**	[1,2]	[1,5] - [1,2]
Scenario 11	[1,2]	**[1,1]**	[1,5] - [1,2]
Scenario 12	[1,2]	**[1,2]**	[1,5] - [1,2]
Scenario 13	[1,2]	**[1,3]**	[1,5] - [1,2]
Scenario 14	[1,2]	**[1,4]**	[1,5] - [1,2]
Scenario 15	[1,2]	**[1,5]**	[1,5] - [1,2]
Scenario 16	[1,2]	**[1,6]**	[1,5] - [1,2]
Scenario 17	[1,2]	**[1,7]**	[1,5] - [1,2]
Scenario 18	[1,2]	**[1,8]**	[1,5] - [1,2]
Scenario 19	[1,2]	**[1,9]**	[1,5] - [1,2]
Scenario 20	[1,2]	**[1,10]**	[1,5] - [1,2]
Scenario 21	[1,2]	[1,2]	**[1,2] - [1,2]**
Scenario 22	[1,2]	[1,2]	**[1,3] - [1,2]**
Scenario 23	[1,2]	[1,2]	**[1,4] - [1,2]**
Scenario 24	[1,2]	[1,2]	**[1,5] - [1,2]**
Scenario 25	[1,2]	[1,2]	**[1,6] - [1,2]**
Scenario 26	[1,2]	[1,2]	**[1,7] - [1,2]**
Scenario 27	[1,2]	[1,2]	**[1,8] - [1,2]**
Scenario 28	[1,2]	[1,2]	**[1,9] - [1,2]**
Scenario 29	[1,2]	[1,2]	**[1,10] - [1,2]**
Scenario 30	[1,2]	[1,2]	**[1,11] - [1,2]**

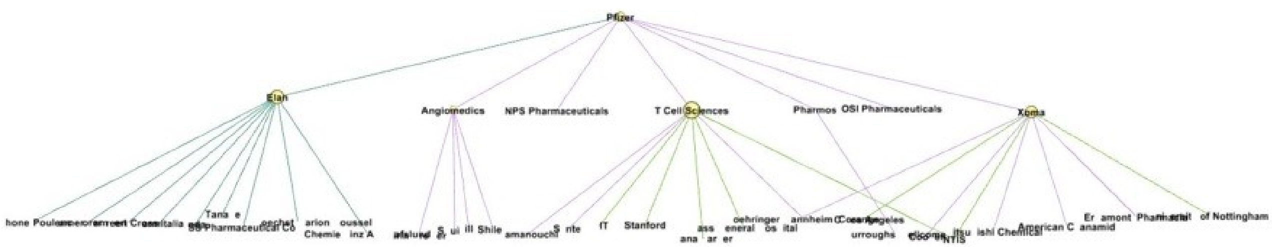
APPENDIX - A

Hierarchical Network Example

1 – Level



2 - Level



3 - Level

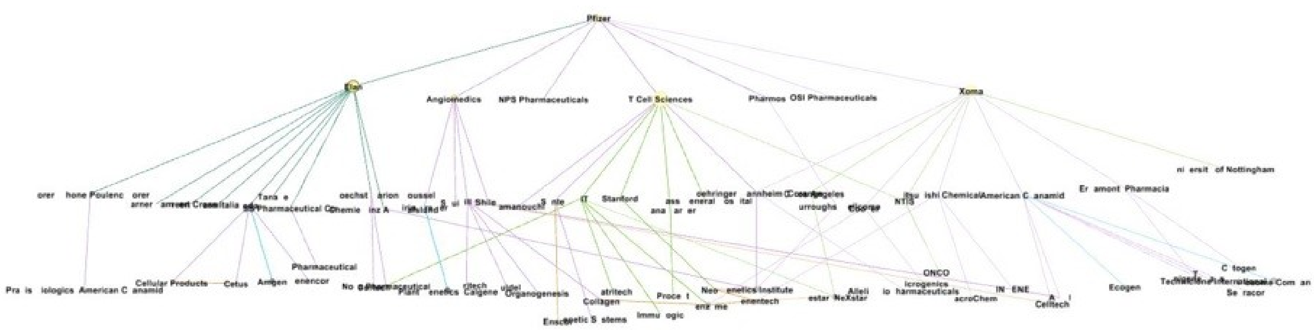


Figure A1 Hierarchical Subnetwork example - Pfizer

Scaling of Scale Free Networks for Simulated Scenarios

This appendix includes figures that highlight the emergence of scale free property for selected simulation scenarios. Since the emergence of scale free property shows slight variation, we have picked scenarios 1, 10, 20 and 30 as representative scenarios for illustration purposes.

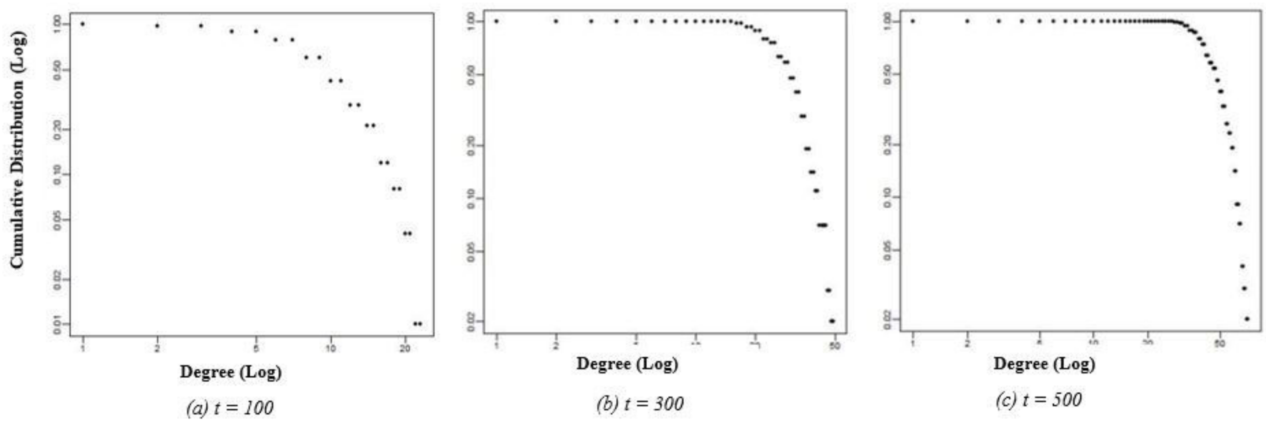


Figure A2 Scale Free Property: Scenario 1

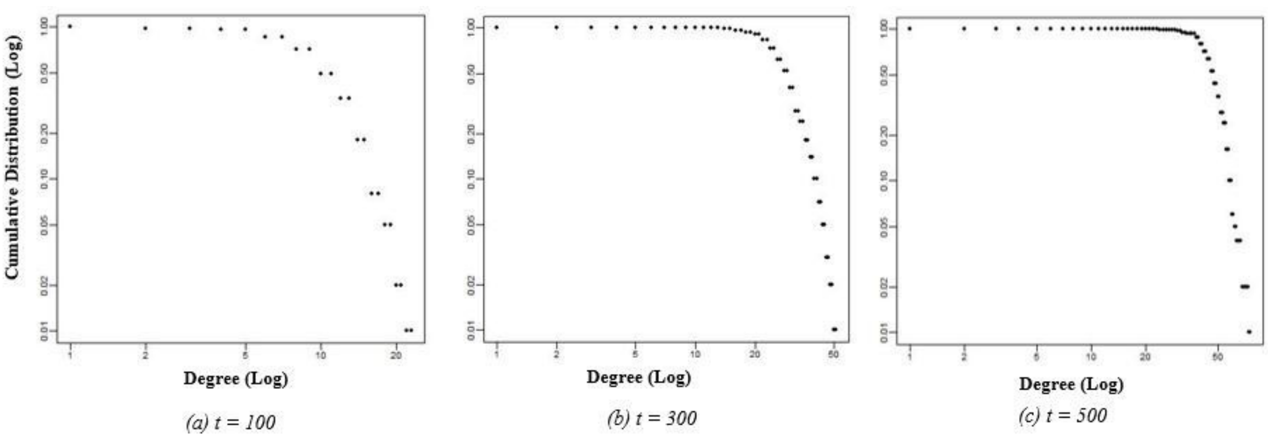
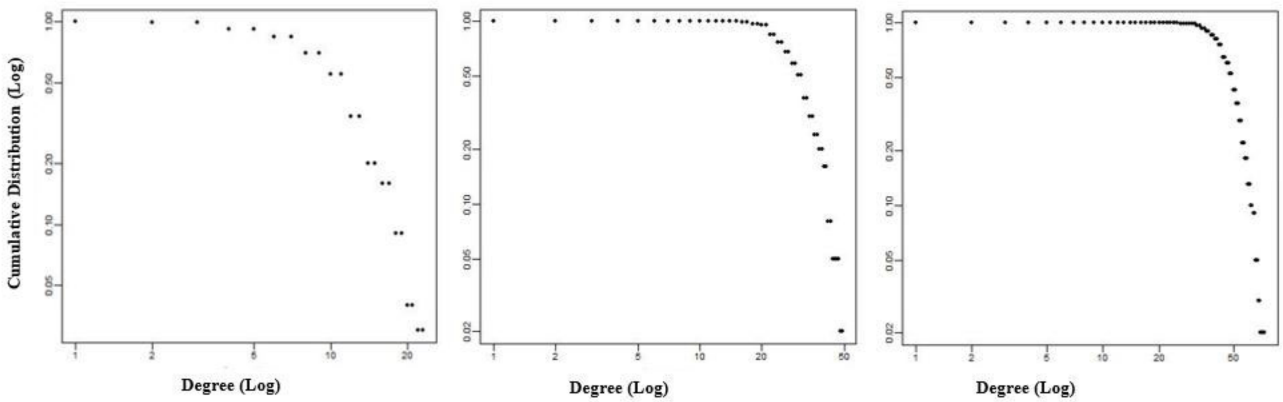
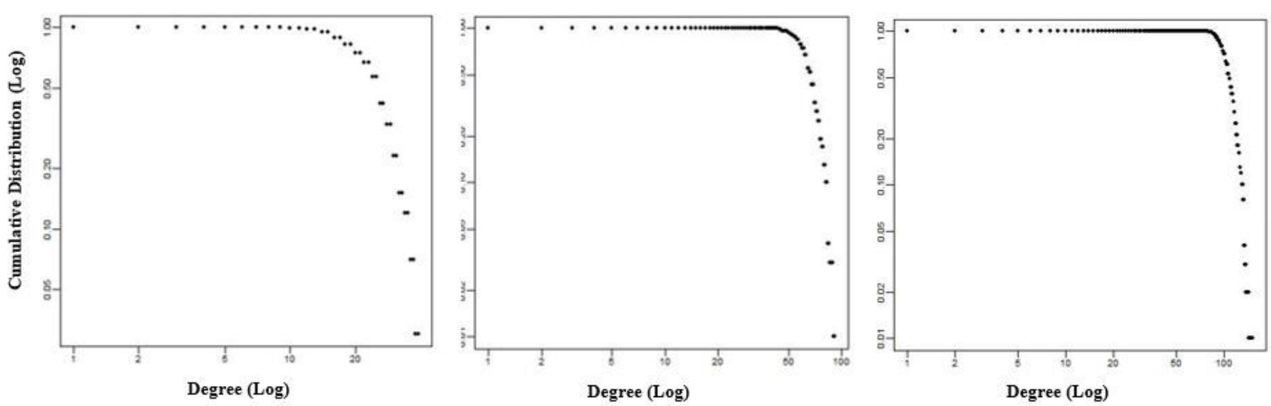


Figure A3 Scale Free Property: Scenario 10]



(a) $t = 100$ (b) $t = 300$ (c) $t = 500$

Figure A4 Scale Free Property: Scenario 20



(a) $t = 100$ (b) $t = 300$ (c) $t = 500$

Figure A5 Scale Free Property: Scenario 30

Scaling of Clustering with Degree For Simulated Scenarios

This appendix includes figures that highlight the scaling property for selected simulation scenarios. For each factor, node fitness, edge fitness and rate of growth, we have included 3 scenarios that are representative: Node Fitness - Scenarios 1, 5 and 10; Edge Fitness - Scenarios 11, 15 and 20; Rate of Partnership Growth - Scenarios 21, 25 and 30.

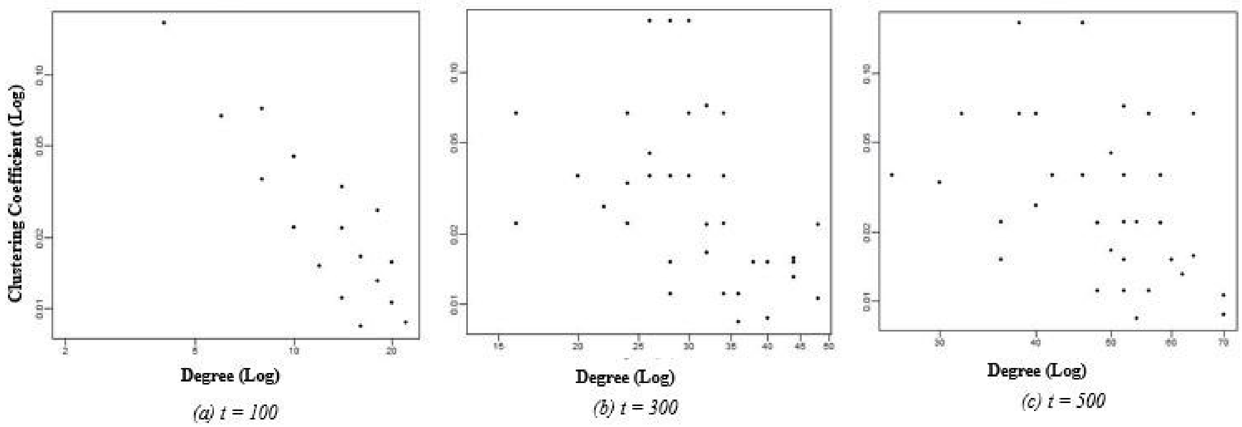


Figure A6 Hierarchical Property: Scenario 1

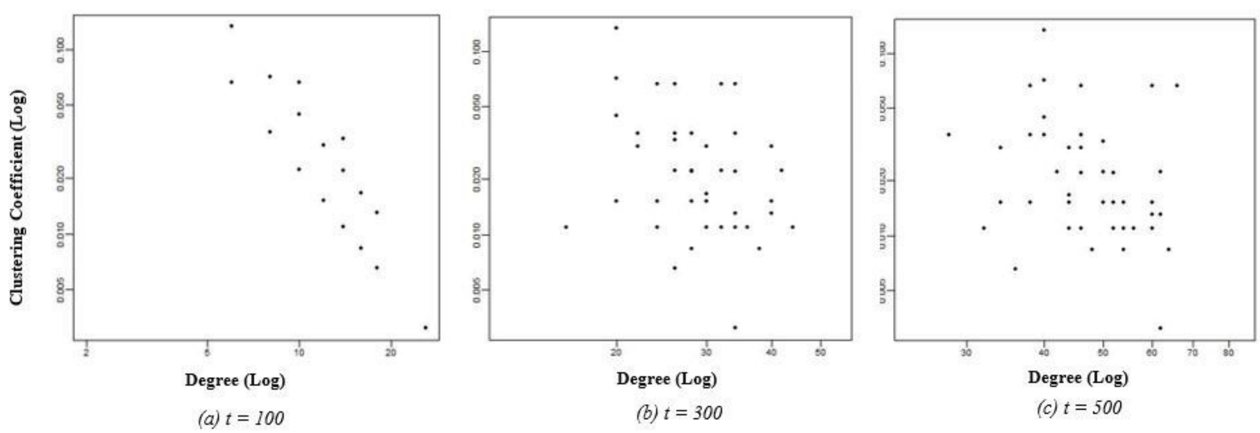


Figure A7 Hierarchical Property: Scenario 5

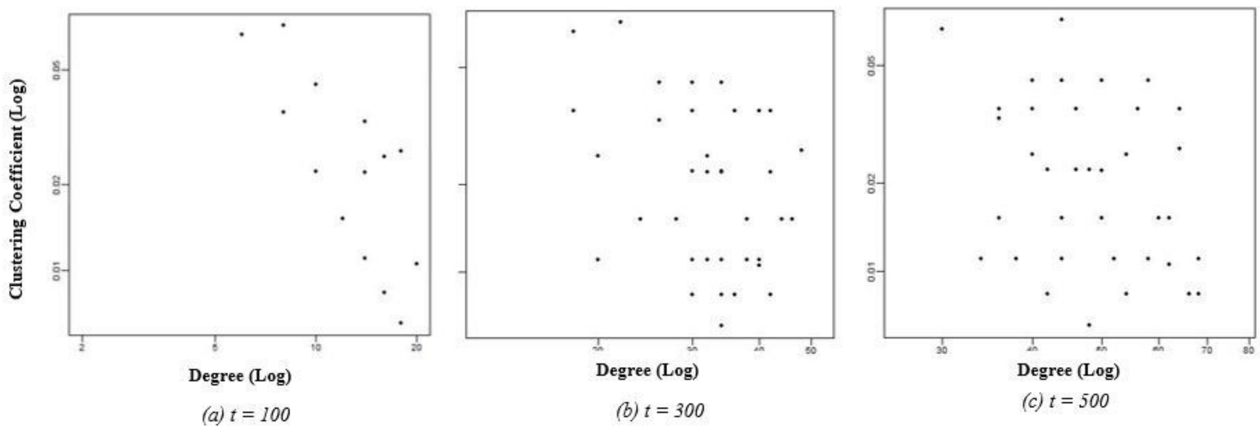


Figure A8 Hierarchical Property: Scenario 10

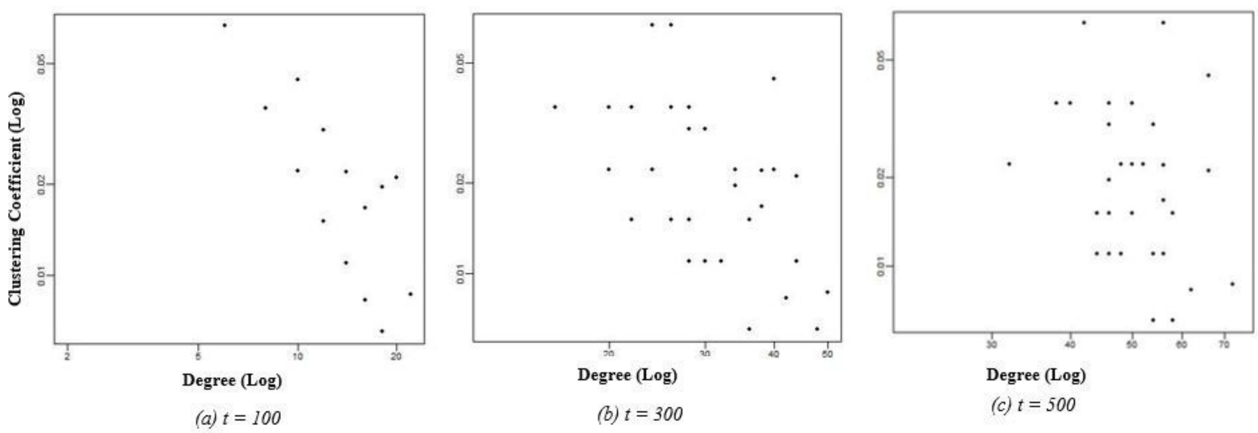


Figure A9 Hierarchical Property: Scenario 11

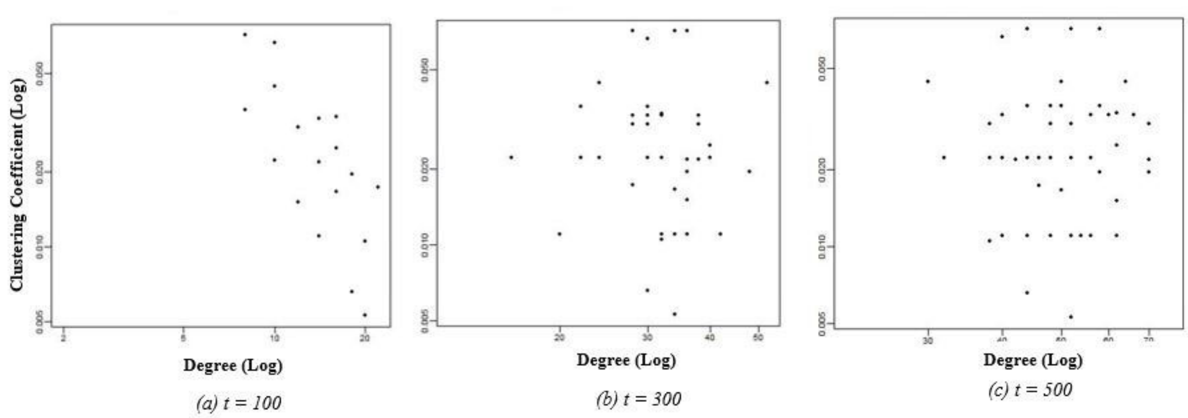


Figure A10 Hierarchical Property: Scenario 15

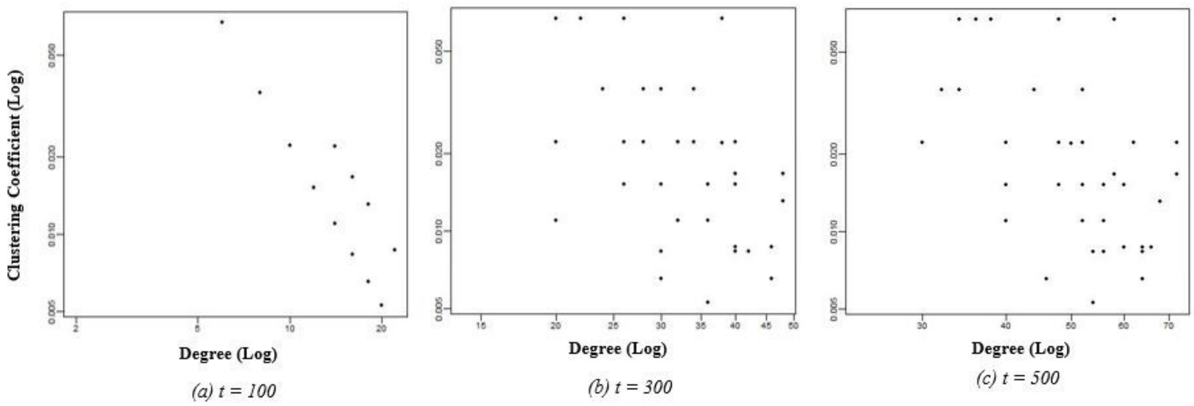


Figure A11 Hierarchical Property: Scenario 20

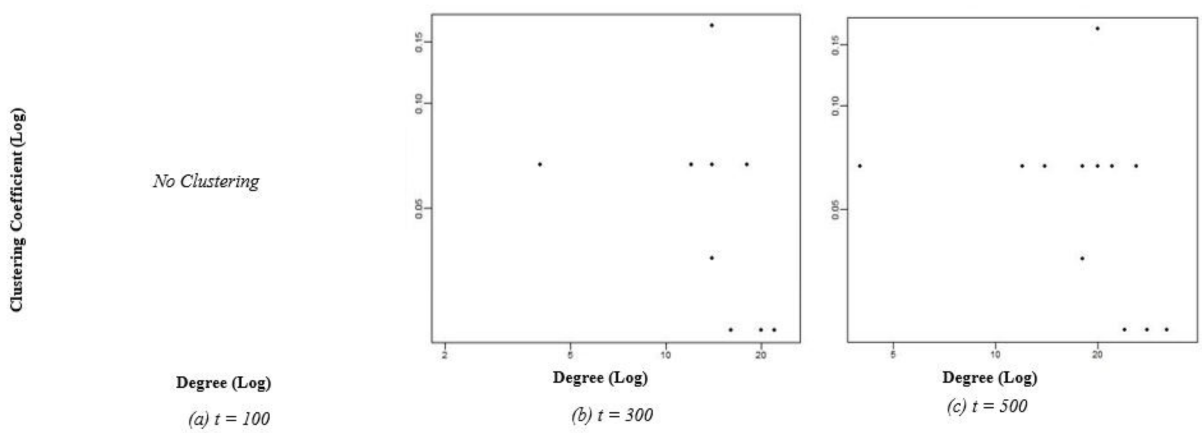


Figure A12 Hierarchical Property: Scenario 21

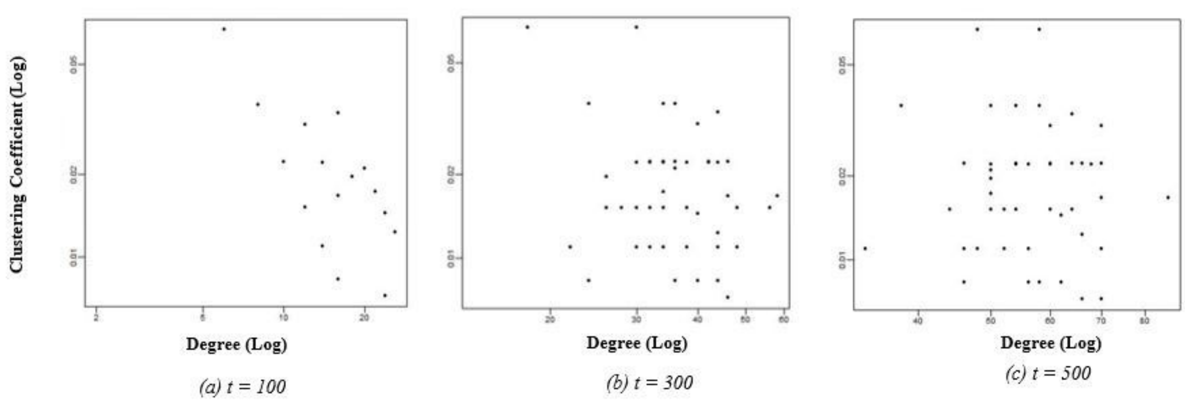


Figure A13 Hierarchical Property: Scenario 25

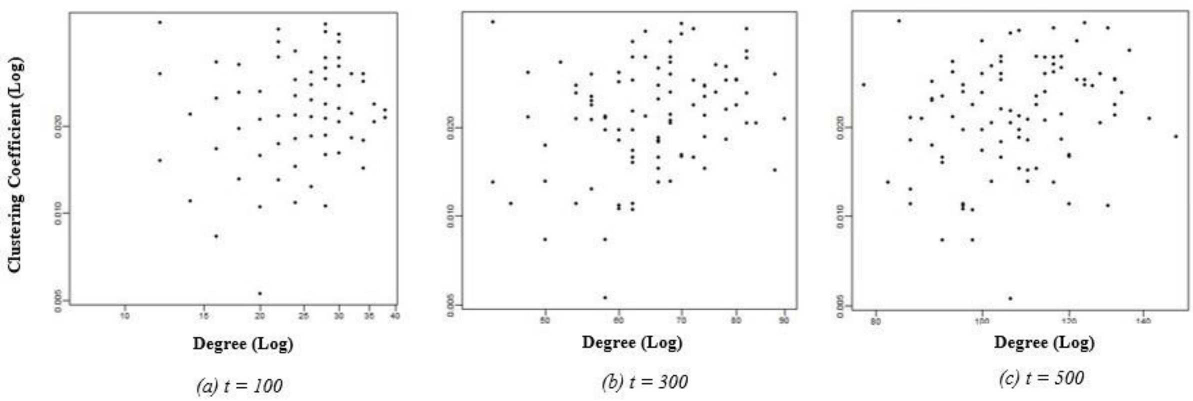


Figure A14 Hierarchical Property: Scenario 30

3. IMPLICATIONS OF PRIORITY ACCESS IN MARKETS WITH EXPERTS: EVIDENCE FROM ONLINE MARKETPLACE LENDING

3.1 Introduction

Effective design and governance of platform markets require all the actors in the platforms, including platform owners and other complementary market players on both sides, to align towards a specific value proposition (Adner, 2017; McIntyre & Srinivasan, 2017; Tiwana *et al.*, 2010). To realize their desired objectives, platform owners must ascertain appropriate economic mechanisms and design rules that can facilitate value creation activities by market players, which in turn, can increase the attractiveness of the platform (Anderson Jr *et al.*, 2014; Baldwin & Woodard, 2009; Dushnitsky *et al.*, 2020; Helfat & Raubitschek, 2018). However, platform owners' design approach to facilitating the intended value proposition often faces significant challenges. Unlike traditional firms, platform owners facilitate transactions between two sides of the market. Inevitably, this role involves market designing activities including the configuration of economic incentives (Boudreau & Hagiu, 2009; Teece, 2018; Yoo *et al.*, 2010; Zhu & Liu, 2018). This is prominently noticed when platform owners have differing objectives or conflict of interest with either one or both sides of the complementors (Casadesus-Masanell & Hałaburda, 2014; Evans & Schmalensee, 2013).

Whenever the design of markets is a possibility, it is pertinent to ask: "how the design of marketplaces influences the functioning of markets" (Roth, 2018). An emerging corpus of literature suggests that platform owners frequently use one such design mechanism, platform market access—their ability to control platform market access by granting or restricting access to certain complementors in either side of the market—to orchestrate economic incentives (Boudreau, 2010; Tiwana, 2013). On one hand, platform owners use their ability to leverage platform market access to spur positive externalities such as network effects or complementarity across products and services to attract more players to the markets and "grow the pie" (Eisenmann *et al.*, 2006; McIntyre & Srinivasan, 2017; Rysman, 2009). On the other hand, platforms can restrict access to the market to manage the desired quality level and attract the right type of complementary players to the platform (Boudreau, 2012; Casadesus-Masanell & Hałaburda, 2014; Halaburda *et al.*, 2018).

This is particularly important as platform owners often resort to targeting supplier/producer side (e.g., drivers in Uber, hosts in Airbnb, workers in TaskRabbit) expansion to cater to increasing demand in the user/consumer side of the market (e.g., passengers in Uber, guests in Airbnb, consumers in TaskRabbit). It is well known that platforms often create an arrangement where they "can affect the volume of transactions by charging more to one side of the market and reducing the price paid by the other side by an equal amount" (Rochet & Tirole, 2006). As these marketplaces grow, platform owners needed additional mechanisms beyond merely subsidizing one side at the expense of the other. Priority access is considered an important design mechanism that helps to control platform access and shape complementors' activities (Cohen, 2017; Evans, 2012; Tiwana, 2013). For instance, Chinese ride-sharing giant Didi Chuxing has made exclusive deals with the local taxi companies to increase the supplier-side availability of the marketplace (Tech Crunch, 2018). Similar interventions by platform owners are widely observed in diverse industrial marketplaces such as hospitality (e.g., Airbnb agreements with boutique hotel and hybrid condo developers), real estate (e.g., Zillow offering partnerships to homebuilders), credit cards (e.g, American Express exclusive offers to Delta Airlines), microblogging (e.g., Twitter's exclusive partnership with media companies), video game consoles (e.g., Microsoft Xbox's partnership with Unity Technologies, provider of the Unity multi-platform game engine to augment the supply of games and identify synergies) and in myriad others.

Incentive design concerns are compounded by the fact that platform owners commonly intermediate the exchanges between both sides of the market in environments where players have heterogeneous levels of expertise in the marketplace (Vallee & Zeng, 2019). Scholars have debated the extent to which expertise matters and whether it compares favorably or unfavorably to the collective intelligence of the crowds (Greenstein & Zhu, 2018; Mollick & Nanda, 2016). Beyond the wisdom of the crowd vs. expert evaluation trade-off analysis, much less is known about how the congregation of experts and non-experts in platforms impacts the performance of the relevant complementors. This is especially important given that platform owners may use governance tools available to them, such as their ability to control market access, to design specific platform markets, which in turn, may lead to tipping the balance in favor of some complementors depending on their level of expertise.

As a first step, we address this gap by examining how the interplay between priority access provided by the platform and the level of expertise of supplier side of the market impacts complementor performance outcomes. Specifically, we study the consequences of platform owners providing deeply resourced suppliers with priority access relative to other same side suppliers. By linking the literature on expertise with the emerging research on platform-based organizations, we argue that platform owners' ability to decide on the platform market access plays a crucial role in the value proposition of complementors. Platform owners with their balancing act of "growing the pie" while trying to pick the "right" complementors, may unintentionally create negative spillovers on the performance and effectiveness of some market participants. We contend that when priority access is given to the experts on the supplier side, it can create lemon-type markets for the non-experts on the same side even when they have similar information about the demand side complementors. Building on this insight, we postulate that the level of priority access (how much of the demand-side is allocated for the prioritized group of suppliers) and cream skimming in the priority market (how much of the demand-side allocated for the prioritized group of suppliers is cream skimmed by them) will amplify the negative spillover effect observed.

To empirically test our predictions, we exploit a novel policy decision with a unique data set from Lending Club (LC), the leading online peer-to-peer (P2P) lending platform in the U.S. that matches potential borrowers with lenders willing to provide credit. P2P lending platforms were established with an initial focus on small, individual, retail investors seeking to participate in the personal loans market. Initially, all loans on the platform were fractionally funded, which meant that each loan would be funded by multiple investors, with each investor usually investing between \$50 and \$250. This approach was ideal for retail investors, who generally had smaller amounts of investing capital, seeking to diversify their portfolios. However, the platform's growth has required the pursuit of larger corporate lenders to sustain the borrower demand and remove the bottlenecks faced in fulfilling the demand side. This led to the introduction of the priority institutional loans market in 2012.

In our study, we exploit Lending Club's decision in October, 2012 to create a whole loans market and implement a randomized process of allocating loans to institutional investors. Loans assigned to the institutional investors would be funded wholly by one investor. Before the implementation of this market, institutional investors that wanted to participate had to do so in the fractional market. This priority access to institutional expert investors in the whole loan market, and the varying levels of it observed over time, offers a unique opportunity to capture the negative spillover effects of this allocation on the performance among the retail investors. We examine the spillover impact of the platform's design choice to provide priority access to the institutional investors. We find that, despite significant evidence that the platform does initially randomly allocate loans, the level of priority access set by the platform has significant negative impact on the returns of the fractional investors. Furthermore, we find that cream skimming by institutional investors also plays an important role in driving same side negative outcomes for non-institutional market participants. Finally, we conclude that more risk-averse fractional investors likely are more strongly impacted by this effect.

This study makes several research contributions. First, we respond to recent calls for researchers to give more attention to mechanism and incentive design issues prompted by the digital technology and mobile telecommunications revolutions in platform-based organizations (Adner, 2017; Adner *et al.*, 2018; Constantinides *et al.*, 2018; McIntyre & Srinivasan, 2017; Teece, 2018; Tiwana, 2013). We extend the literature that suggests platform owners play a governance and regulatory role by emphasizing how platform market access prioritized to some complementors is used as a key design tool (Boudreau & Hagiu, 2009; Evans & Schmalensee, 2013). Understanding the use of priority platform market access provides insights into the nuances of value appropriation among complementors. We demonstrate how lemon problems can creep into the platform markets to some of the complementors on the same side of the market when design changes are undertaken by the platform owner to "grow the pie" and serve the demand-side customers. Second, a growing stream of literature discusses the challenges of cohabitation of more informed experts and less informed amateur crowd in the platform markets (Boudreau, 2018; Greenstein & Zhu, 2018; Mollick & Nanda, 2016; Vallee & Zeng, 2019). We undertake an initial effort to mitigate an important gap in this literature: the interaction effect of platform design/governance mechanisms and the level of expertise of the complementors with their performance. Our study complements the prior research that focus on the performance differentials between experts and non-experts by showing how negative spillover effects can emerge *within* supply-side complementors prompted by platform design choices. Third, we link new platform market literature with the classic corporate strategy research focusing on governance choices. Platform owners unlike traditional firms do not have fully formed contractual relationships with the complementary market players. Instead they are often associated with the platform through loosely connected alignment structures with possible mutual benefits (Adner, 2017; Boudreau, 2017). Therefore, it is essential to expand the corporate strategy research beyond conventional means such as hierarchical control or contractual lens to understand platform-based organizations (Adner *et al.*, 2018; Barach *et al.*, 2019; Chu & Wu, 2019; Helfat & Raubitschek, 2018). In our view, the governance and design perspective illustrated in this paper is a useful addition to the conventional corporate strategy research on both within and interfirm relationships that show how governance choices affect the organizational and transactional performance.

3.2 Theoretical Background

The successful orchestration of a value proposition in platform depends on how effectively platform owners can design the market and govern the configuration of economic incentives for the complementors (Adner, 2017; Boudreau & Hagiu, 2009; McIntyre & Srinivasan, 2017; Tiwana, 2013). Prior to the digital revolution, a vast majority of the mechanism and market design discussions were restricted to markets such as school choice, marriage matching and kidney exchange (Goldfarb & Tucker, 2019; Roth, 2018). However, the intermediating nature of digital platforms, as well as the possibilities provided by a vast array of business models, has created the opportunities for platform owners to become involved in market design in their platforms. Vulkan *et al.* (2013) summarizes this distinction as follows:

"Economists often look at markets as given, trying to make predictions about who will do what and what will happen in these markets. Market design, in contrast, does not take markets as given; instead, it combines insights from economic and game theory together with common sense and lessons learned from empirical work and experimental analysis to aid in the design and implementation of actual markets."

Platform owners need to align appropriate set of actors in the platform to realize a core value proposition (Adner, 2017; Teece, 2018). Recent research has shown that platform governance is used to shape the complementor activities by effective market design (Boudreau, 2017; Dushnitsky *et al.*, 2020; Rochet & Tirole, 2003; Yoo *et al.*, 2010). This is often steered by the careful planning and deployment of governance tools such as control mechanisms, decision rights and ownership structure (Cumming *et al.*, 2019; Tiwana, 2013).

Among the possible platform governance instruments for market design, control of platform market access has gained particular attention among scholars and practitioners alike (Boudreau, 2017; Evans & Schmalensee, 2013; Parker & Van Alstyne, 2018; Stigler Committee on Digital Platforms, 2019). Platform market access refers to platform owners' ability to grant or restrict access to certain complementors in either side of the market. From a legal and antitrust perspective, it has created discussions about the possible challenges resulting from a lack of separation between a platform and the commerce conducted on the platform. Moreover, there is significant debate regarding whether platforms should be excused for the substantial amount of consumer surplus they create or penalized for not adhering to the common carriage principle of "nondiscriminatory public access" and the expected "indifference to the nature of the goods carried" (Cohen, 2017; Evans & Schmalensee, 2013; Stigler Committee on Digital Platforms, 2019).

From a more strategic and organizational perspective, the control of platform market access by platform owners generates an important series of inquiries about how the platforms regulate value creation and appropriation process by controlling the lever of platform market access. Extant research has shown nuances on the impact that economic incentives of the complementors have on marketplaces. More generally, platform owners often end up either leveraging their ability to control platform market access to achieve desirable outcomes and "grow the pie" (e.g., network externalities, offering complementarity products and services) or using it to restrict access to curate the marketplace (e.g, monitor the quality of complementors, attract right type of complementors) (Boudreau, 2012; Casadesus-Masanell & Hałaburda, 2014; Eisenmann *et al.*, 2006). Platform owners also frequently use the control of market access as a lever of influence to achieve outcomes favorable to them. For instance, Amazon often retain rights to sell wholesale, pricing and customer data for providing third party sellers access to the marketplace. In a related vein, it is widely observed that market access can be used as a tool of competitive actions by the platforms. For example, Facebook denied API access to perceived competitors such as Vine, a Twitter-owned feature that let users create short videos when they released a competing product, leading Vine to be defunct. Platform owners' regular use of "market access for data" approach—free or discounted access to the platform in lieu of user/ consumer side of the market sharing their data—is also heavily discussed and scrutinized.

Among various methods to control platform market access, priority access has received prominent attention as a salient design tool for platform governance (Evans & Schmalensee, 2013; Stigler Committee on Digital Platforms, 2019). Priority access refers to platform owners' ability to provide some subset of complementors with preferential treatment in terms of access to the marketplace. Priority access can be used by platform owner to shore up one side of the market with selected complementors. More importantly, through priority access, the platform owner might be able solve demand-side bottlenecks by providing preferential treatment for a portion of the supplier-side complementors. For instance, when credible borrower demand in the online lending peer to peer (P2P) industry is higher than the lenders' supply of money, a lending platform owner may resort to giving priority access to expert institutional investors such as investment banks, pension funds, commercial trusts, endowment funds, hedge funds, and private equity investors.

Although existing studies have offered substantial insights into the implications of platform market access, how priority access generates spillover effects on the same side complementors without priority access remains unclear. Given the heterogeneous levels of expertise in complementors on the supplier side of the marketplace, it is vital to understand how the platform governance choices such as priority access can bring performance differential due to the interplay between the two. For example, what happens to the crowd retail investors when institutional investors get the priority access in online P2P lending? This leads to a discussion on the institutional details of our research context of online P2P lending platforms below.

3.3 Institutional Background

In this paper, we focus on Lending Club (LC), which is the largest P2P lender in the USA. In 2018, LC originated $20 billion in personal loans accounting for 45% of the online P2P market share (Forbes, 2017). These platforms have experienced tremendous growth since their nascent year with loan volume in the US-based online P2P lending platforms growing at an average of 84% since 2007 (PWC, 2015). By mid-2017, online P2P lending platforms accounted for about one third of all the personal loans originated in the US market and estimated to grow to $150 billion in 2025 in the United States (PWC, 2015; Trans Union, 2017). Online P2P lending provides a suitable avenue to study the interplay between platform governance and the level of expertise of the complementors. For the supplier-side lenders, success in the platform depends on their ability to pick the right borrowers with the accurate valuation on the future payoff. Financial market literature suggests the importance of expertise for investors originating from their institutional knowledge (Shleifer & Summers, 1990; Vallee & Zeng, 2019). It provides opportunities to study the implications of having experts and non-experts together in a designed marketplace such as online P2P lending and to understand how platform governance can create differential performance implications to them.

On LC, potential borrowers submit their loan applications and provide details regarding their requested loan amount, loan term, and loan purpose. If the borrower passes LC’s initial screening phase, LC subsequently assigns a grade, ranging from A1 for the least risky borrowers to G5 for the riskiest borrowers. A borrower’s grade is assigned based upon the platform’s proprietary algorithm which predicts a borrower’s propensity to default and determines the interest rate they are assigned by the platform. If the borrower accepts the platform's assigned interest rate, the potential loan is then listed and investors can choose whether they wish to participate in the funding of this loan.

Prior to October 2012, all potential loans were listed in the retail loans market. In this market, investors could choose to provide fractional funding for a loan with a minimum amount of $25. Since LC was initially designed to provide an opportunity for small retail investors to participate in the credit market, the fractional loans market was ideal for retail investors, who generally had less investing capital, seeking to diversify their portfolios. However, the sustainability of these platforms is dependent upon the expansion of both the borrower and lender pools, which has necessitated that P2P platforms aggressively pursue new users on both sides of the market. While P2P lending platforms were established with an initial focus on retail investors seeking to participate in the personal loans market, their growth has required the pursuit of larger institutional lenders to sustain the borrower demand. However, the potential growth provided by institutional investors incentivized LC to pursue them and provide priority access in the P2P marketplace.

To fulfill the demands associated with incorporating institutional investors, LC introduced the priority institutional loans market in October, 2012 (LC refers to this market as the whole loans market). While the fractional retail loans market was ideal for investors with limited capital, institutional investors had significantly more investing capital, and their diversification needs could be met by purchasing a large portfolio of whole loans instead of small fractions of each loan. Lending Club stated the following regarding the introduction of the whole loans market in 2012:

"To accommodate these requests and while insuring that all investors continue to have access to a large number of loans of equal quality, a randomized subset of loans by grade will be available to purchase as a whole loan (i.e. not in $25 increments) only for a brief time period (12 hours), while all other loans will be immediately available for fractional purchase. If the loans are not purchased as whole loans in the specified time period, they will become available for purchase in the standard, fractional manner."

Therefore, after the introduction of the priority institutional loan market in October 2012, the platform stated that each listing was randomly allocated to either the retail loan market or the newly created priority institutional loan market. While all investors could participate in the fractional retail loan market, the priority institutional loan market was comprised of banks, asset managers, insurance companies, hedge funds, and other large non-bank investors.[10] The platform seeks to mitigate any negative impact on retail investors that this privileged access would have by doing the following: 1) randomly assign loans to each market within grade and 2) provide the retail investors the opportunity to fund loans that are not funded by institutional investors if they choose. To emphasize the continued importance of the retail investors despite the changes, LC stated the following at the introduction of the priority institutional loan market:

"The design of our platform emphasizes how important retail investors are to LendingClub. We see our retail investors as a key component of our diverse marketplace strategy. Retail investors are—and will always be—the heart of the LendingClub marketplace." (Lending Club, 2012)

LC suggested that the randomization serves to protect the retail investors from any potential harm that may come from the priority access given to the institutional investors. Therefore, LC argued that this design change will retain the successful aspects of the platform while keeping retail investors at the "heart", while also bringing necessary changes to move to the next step of platform growth. This novel setting helps us to study what happens to different type of supplier-side investors when platform changes the way the loan allocation process works—randomly assign the loans at the grade level to the priority institutional loan market—and, subsequently, the rejected loans by the institutional investors sent into the retail market.

3.4 Hypotheses

3.4.1 Interplay between Priority Access and Expertise

Expertise of the complementors vary within a platform. The involvement of expert complementors often brings much needed credibility to nascent platforms. The presence of expert complementors also attracts additional complementors to both the same and the opposite side of the market, facilitating platform owners' overall pursuit to "grow the pie". Thus, as marketplaces grow, platform owners, realizing the importance of expert complementors to the overall value proposition of the platform, tend to cultivate their participation. As said, in online P2P platforms, expert

[10] LC, in their 2019 10-K filings, state the following description of their Whole Loan Product: "Certain institutional investors, such as banks, asset managers, insurance companies, hedge funds and other large non-bank investors, seek to hold whole loans on their balance sheets. To meet this need, we sell whole loans to these investors through loan purchase and sale agreements."

institutional investors responding to the FinTech revolution have provided a vital supply-side solution to platform owners to satisfy increasing customer demand.

Evidence in the literature is mixed when comparing the expert performance to the wisdom of the crowds. Studies have laid out the contingencies under which wisdom of the crowd may supersede expert decision making. On a general level, scholars have noticed that when crowd complementors have the ability to coordinate through an appropriate set of authority figures, the benefits attributed to the wisdom of the crowd can be realized (e.g., lead developers in large R software packages, moderators in successful Reddit forums, dedicated curators in niche news platforms such as Hacker News and editors of Wikipedia) (Jeppesen & Frederiksen, 2006; Mollick & Nanda, 2016; Poetz & Schreier, 2012). The crowd-based approach works better when the small contributions of each crowd complementor can be aggregated towards a large goal or outcomes.

In online P2P lending platforms, mitigating circumstances for crowd retail investors to perform better than the expert institutional investors are scant. First, retail investors don't coordinate with each other or have an authority figure within them to gain insights from each other. Instead they use their own independent judgement in deciding whether to fund a potential borrower. Therefore, the benefits often realized in successful crowd-based platforms such as Wikipedia are limited. Second, both the institutional and crowd investors use the same set of information in making decisions. Experts, when presented with the same information, often utilize it more effectively than the crowd (Kim & Viswanathan, 2018). Prior research has also shown that, in the financial markets, institutional investors act as "smart money" and perform better through their superior selection and screening ability (Shleifer & Summers, 1990). Taken together, we posit that expert institutional investors, when utilization of their expertise is plausible, will generate better payoffs compared to the crowd retail investors.

To examine the impact of priority access accorded to the expert institutional investors by the platform policy change, we follow the past research in both platforms and financial markets, focusing on the incentives on FinTech platforms such as online P2P lending. We propose that priority access granted to the experts benefits them even after accounting for their expertise. Selectively given market access could reduce competitive intensity among complementors and potentially increase the attractiveness of the platform (Casadesus-Masanell & Hałaburda, 2014; Halaburda *et al.*, 2018). Consider the introduction of priority market for the expert institutional investors in online P2P lending discussed previously. For institutional investors, the prioritized allocation of a subset of loan listings as priority institutional loans provides a space to search for the right borrowers without getting contaminated by the retail crowd investors. This is consistent with the finance research that suggests that when there is reduced competition among the investors it can assist the institutional investors to do quick search and fast-track the process of information acquisition about the investment landscape. For example, Biais *et al.* (2015) indicated that the delays in execution due to search concerns account for about one-third of total costs for institutional investors. Providing priority access to the expert institutional investors help them to alleviate potential “arms race” they may otherwise find themselves in (Glode *et al.*, 2012). Taken together, priority access provides an avenue for institutional investors to face reduced intensity of competition among potential investors, leading to the increase in speed and efficiency with which they can determine the value of potential loan listings.

For the retail investors, priority access given to the institutional investors accentuates their disadvantage even after accounting for their lack of expertise. Priority access provided to the expert institutional investors shrinks the available loan listings to them even though the overall pool is randomized before institutional investors getting the priority access. It forces the retail investors to search for suitable loans within vanishingly small set loan listings, increasing an “arms race” within competitive pool of lenders. Thus, we hypothesize:

Hypothesis 1: *The higher the level of priority access granted to experts, the intensity of the negative same-side spillover effects on the crowd increases in online P2P lending platform.*

3.4.2 Cream Skimming and the Emergence of Lemon Market

A potential source of market failure in the designed markets is that the competitive pressure may lead the supplier-side of the market to aggressively focus on the high-demand customers (the "cream") and leave the low- demand ones underserved or unserved (Kahn, 1988; Laffont & Tirole, 1993). This is noticeably prevalent when the supplier-side has limited incentives or obligations to satisfy different types of customers (Laffont & Tirole, 1990). Relatedly, prior research has noted that supplier-side sellers may opt to ignore the skimmed high cost, low profit market segments even if the long run marginal cost is lower than the price charged at least for two reasons. First, opportunity cost of serving the skimmed market may not be worth to spend the firm’s resources and investments on, leading firms to look elsewhere where the benefits are higher. Secondly, even if the firm opts to serve this segment of the market, they might make themselves vulnerable to other competitive players including potential new entrants who may cream skim the market and be able to peel off the customers. Prior literature spanning in wide range of contexts, often in the regulatory intensive industries, has observed the implications of cream skimming (e.g., financial markets, health insurance, education, telecommunications, cable television, utilities and airlines) and how it can impact the effectiveness of the markets.

Cream skimming challenges are acutely observed in the financial markets as the informed investors or the investors who has acquired costly information (Bolton *et al.*, 2016; Vallee & Zeng, 2019) devote resources to accurately gauge the future payoff, leading to adverse selection problems for the rest (Fishman & Parker, 2015). To avoid the unraveling of the market, institutional safeguards are put forward with hard regulations such as the SEC protocols that requires concerning parties to disclose all significant information to all investors, or in the extreme cases banning any form on insider trading outright. Similarly, softer rules such as the speed limits in which the algorithmic transactions and trades are designed to alleviate the concerns arising from cream skimming and ensure the “level-playing” field for all.

Platform owners often play the role of internal regulators. Platform market literature has revealed that the design rules are effective for the platform owners to execute a regulatory apparatus. For example, to mitigate the cream skimming by the drivers, ride-sharing platform Uber does not disclose the passenger’s destination to the drivers until after the driver has picked up the passenger. This helps to alleviate drivers from waiting for longer rides or preferred destinations that can lead towards longer waiting times for the passenger (Romanyuk & Smolin, 2019). This also avoids scenarios where expert drivers strategically pick good rides, leaving relatively bad ones to more inexperienced drivers. However, when there is a conflict of interest between the platform and some set of complementors, it can create challenging contingencies in the configuration of economic incentives regarding cream skimming.

Priority access to the experts in platforms leads to the unmet demand from the sample allocated to the experts directed towards the rest of the pool. In our online P2P lending context, if institutional loan market listings are not funded by the institutional investors, they are reassigned to the retail market. Despite initially providing priority access to grow the pie, platform owners reassign the loans if they are not picked by the institutional investors to the retail market so that the demand doesn’t go unfulfilled. As indicated before, expert institutional investors, when they can utilize their expertise, generate better payoff compared to the crowd retail investors. While there is no obligation from the retail investors to fund the loans recycled from the institutional loan market, a potential issue arises if these recycled loans are systematically inferior and retail investors do not incorporate this information into their decision-making process. When benefits of wisdom of the crowd is not outweighed by the expert evaluation in online P2P lending, rejected loans by the institutional investors perform worse than those picked by them. With the "back door" policy that sends the rejected institutional loans to the retail market, platform priority access policy generates "lemon" problem for the retail investors, leading to the issue of adverse selection. Thus, we hypothesize:

Hypothesis 2a: *Cream skimming by the experts has same-side negative spillover effects on the payoff of the crowd in online P2P lending platform.*

Hypothesis 2b: *An increased exposure of the crowd to the residuals of cream skimming by the experts leads to higher same-side negative spillover effects in online P2P lending platform.*

3.5 Data and Methods

3.5.1 Research Design

To study our proposed hypotheses, we focus on online P2P lending and exploit the introduction of the separate priority institutional loans market by LC. Specifically, we utilize periodic variation in the level of priority access LC assigns to the institutional investors. Figure 3.1 Panel I displays the monthly variation in the proportion of loans LC initially assigns to the institutional investors. This figure provides suggestive evidence for the temporal variation in priority access granted by the platform. Moreover, Figure 3.1 Panel II provides evidence that proportion of priority access by grade also varies temporally.

Our empirical strategy is to evaluate the conditional expectation of retail loan outcomes on changes in platform assigned privileged access levels. We define the level of priority access as the proportion of the overall loan requests made that are assigned to the priority institutional loans market for a specific week. Figure 3.2 shows the temporal variation in the monthly assignment of privileged access by LC. Figure 3.3 displays the weekly variation in level of priority access for 2013 to 2016. These figures outline the weekly variation in the level of priority access provided by the platform. Moreover, they negate concerns regarding temporal patterns potentially driving the variation we see in priority access set by the platform. In our specifications, we control for year level unobservable factors by including grade by year fixed effects, which we discuss in more detail below.

A critical feature of this design is LC's decision to randomly allocate loans, within grade, to either the institutional investors or the retail investors. This is vital to our design as it alleviates concerns that the platform might be selectively allocating loans of different quality to each market. If LC was not randomizing loan assignment, any evidence relating retail market outcomes to levels of institutional market priority access is potentially a result of the platform's loan allocation decisions. In contrast, we seek to identify the impact of the platform implementing various levels of priority access and not the decision to assign a specific loan to either market. Randomizing allocation creates an ideal situation to isolate the role of priority access on complementor outcomes. Furthermore, the nature of P2P lending is such that the transactional outcomes between borrowers and lenders have clear comparable outcomes—an A1 grade loan that defaulted is inferior to an A1 grade loan that does not default. This highlights another important feature of this data set, providing a unique insight utilizing detailed transactional level data to study platform markets with priority access.

Another advantage of this design is that the platform reassigns the loans rejected by the institutional investors to the retail loan market. This provides a novel setting to study the interplay between priority access and cream skimming by experts. It enables us to empirically examine how the retail market may become contaminated due to cream skimming by the experts and the impact it would have on the performance of non-expert suppliers. The platform's decision to randomize loan assignment is beneficial for this analysis as well since it would have been difficult to isolate the effect of expert cream skimming from the platform's decision to assign better or worse loans to a group of supply-side investors.

3.5.2 Sample Construction

We gather LC data from various sources, obtaining all 3-year term LC loans from October 2012 to August 2016. We combine LC publicly available loan data set and note sales from their SEC filings, with loan information obtained from each loan's information profile on Lendingclub.com.[11] By matching the SEC note sales and the publicly available loan data we are able to ascertain whether loans in the retail loan market were originally provided to the institutional investors and subsequently rejected by the institutional investors.[12] While LC also offers 5-year loans we focus on 3-year loans since we are unable to examine the conclusion of 5-year loans that originated past 2014.

Our full data set consists of 783,999 loans. As previously shown in Figures 3.1 – 3.3, LC is constantly adjusting the level of privileged access which they are providing to the institutional investors. For each loan, we obtain detailed information on each borrowers' credit characteristics including debt-to-income ratio, homeownership status, annual income, FICO score, and the number of open credit lines. We also obtain information on all payments made by each borrower and the timing of these payments. This allows us to calculate the rate of return of each loan. Each loan is classified into one of three categories: retail loans, institutional loans, and loans rejected by institutional investors. Retail loans are those that are randomly assigned to the retail loans market. Institutional loans are loans that are randomly assigned to the priority institutional loans market and funded by the institutional investors. Loans rejected by institutional investors are loans that were initially assigned to the priority institutional loans market but were not funded by the institutional investors. Loans rejected by institutional investors are subsequently recycled into the retail loans market and made available to the retail investors.

To evaluate the hypotheses proposed, we construct a sample of loans that includes all retail loans and loans rejected by institutional investors. We refer to this sample as the *Full Retail Market Sample.* This sample includes all loans that were made available to the retail investors in the retail loan market, whether through initial random allocation or recycled from the priority institutional loan market. Since we are interested in evaluating the platform's assigned level of priority access on retail participants, the *Full Retail Market Sample* is our main sample. There are 378,487 loans in this sample.

[11] By LC publicly available loan data we refer to data readily accessible on the company's website. The loan information profile information provides extra information such as the number of investors for each loan and the exact date the loan was listed and funded.

[12] Prior to October 2014, it is not possible to use the SEC notes sales to identify the rejected institutional loans. This is documented by LC. However, we can determine whether a loan is rejected by the institutional investors prior to October 2014 because the obtained data from the information profile page of each loan provides the number of eventual investors that funded each loan. If a loan was initially assigned to the whole loan market but had a large number of investors we can ascertain that it was rejected by the institutional investors and was converted to a fractional loan.

3.5.3 Variables and Measurement

Dependent Variables.

We employ two dependent variables in our specification: *Default* and *Rate of Return*. *Default* has a value of 1 if the borrower of the loan defaults on their loan obligations, otherwise *Default* is assigned a value of 0. *Rate of Return* is the overall percentage return of a loan.[13] For example, if a borrower with a $5,000 loan made payments totaling $5,500, the *Rate of Return* for that loan would be 10%. These variables are commonly used to evaluate the performance of loans and investors specifically on P2P lending platforms and credit markets in general (See Morse (2015) for review).

Figure 3.4 outlines the relationship between borrower assigned grade, interest rate, and the two outcome variables (*Default* and *Rate of Return*). As expected, riskier borrowers are assigned higher interest rates and have higher default rates. Since the platform assigns grades based on default predictions, this is expected. Figure 3.4 also shows that there is not as much variation in the rate of return, however, the variance of returns is higher when the loans are riskier. This figure also emphasizes the importance of grades in determining loan outcomes.

Explanatory Variables.

To study the various hypotheses prescribed previously, we utilize a collection of explanatory variables. *Proportion Priority Institutional Loans* is the proportion of LC loans that were assigned to the priority institutional loan market in a specific week. This represents the level of priority access provided by the platform and is used to examine hypothesis 1. *Loan Rejected by Institutional Investors* is an indicator variable which has a value of 1 if the loan was initially assigned to the priority institutional loan market but was not funded by the intuitional investors. Recall that if the loans are rejected by the institutional investors (*Loan Rejected by Institutional Investors* has a value of 1, otherwise 0) then these loans are recycled to the retail market and retail investors can fund them if they wish. This variable is used examine hypothesis 2a. *Proportion Loans Rejected by Institutional Investors* is the proportion of loans in the retail market in a specific period that were initially assigned to the priority institutional loan market but not funded by the institutional

[13] Specifically, rate of return is calculated by the following formula: $\frac{Sum\ of\ Payments\ made\ by\ Borrower}{Funded\ Amount}$. We also examine the robustness of our results to measures of rate of return that incorporate the time-value of each payment. Our results are consistent and available upon request.

investors. These loans are reassigned to the retail market. As the *Proportion Loans Rejected by Institutional Investors* increases, the greater the spillover to the retail market resulting from the platform's decision to recycle the loans rejected by the institutional investors. This variable is used to assess hypothesis 2b.

Control Variables.

We also include a collection of variables to control for loan specific attributes that may cause changes in the conditional expectation of the loan's default or rate of return. As portrayed in Figure 3.4, a borrower's platform assigned grade is critical in determining the outcome of the loan. Moreover, Figure 3.1 Panel II shows that the proportion of priority access varies not only temporally, but also temporally by grade. Therefore, we include grade by year fixed effects for each unique combination of year and grade. These controls for the differences associated with the grade and any temporal changes in the outcomes relative to borrower grade. This is especially important given that the platform randomizes by grade. Therefore, when including this control, we can ascertain whether platform assigned priority access impacts the retail market after controlling for the grade. Consistent with prior research, we further include the following borrower/loan specific variables: *Funding Amount, Debt-to-Income, Inquiries in Last 6 Months, Annual Income, Revolving Balance, Interest Rate, Homeownership Status, Current Credit Lines* (Butler *et al.*, 2017; Morse, 2015). Table 3.1 provides a detailed description of each variable. Table 3.2 displays the descriptive statistics and correlation of the variables used in our analysis.

3.6 Results

3.6.1 Platform Random Allocation of Loans

LC explicitly state that they randomize loan access between institutional and retail loan markets, both on their website and in their 10-K filings. The Securities and Exchange Commission (SEC) provides oversight of P2P lending. Therefore, any departure by LC from their stated intention to randomize allocation would result in steep legal ramifications. Nevertheless, to empirically investigate LC's within grade randomization of loan assignment, we regress outcome variables (*Default* and *Rate of Return*) on a binary variable indicating whether a loan was issued to the institutional investors or the retail investors. We refer to this variable as *Priority Institutional Loan* (value equals 1 if the loan is assigned to the priority institutional loan market and 0 if assigned to the retail market). Since the platform randomizes loan assignments within grade and this randomization pattern has a temporal variation, we control for grade by year fixed effects. Controlling for grade by year fixed effects the loan was listed should remove any difference between the loan outcomes in the priority institutional loan market and the retail loan market respectively. Therefore, we include fixed effects for all the unique combinations of borrower grade and year. The specification is as follows:

$$Outcome_i = \beta_0 + \beta_1 * Priority\ Institutional\ Loan_i + \gamma_i + \epsilon_i$$

This specification is estimated at the individual loan level (i). $Priority\ Institutional\ Loan_i$ has a value of 1 if the loan is assigned to the priority institutional loan market and 0 otherwise. γ_i corresponds to the grade by year fixed effect specific to the grade of loan *i* and the year loan *i* was originated. This model estimates the difference in means, after controlling for the specific grade and year of each loan, between the loans initially assigned to the institutional investors and those assigned to the retail investors. If the platform is in fact randomizing the loan assignment, there should not be a difference between outcomes in the retail or institutional loan market and β_1 should not be statistically different from 0.

Columns 1 and 2 of Table 3.3 show the results of this specification and the evidence suggests that the platform is indeed randomizing the assignment. Specifically, we find that the platform's initial assignment of a loan to the priority institutional loan or retail loan market does not have an impact on the conditional expectation of the loan regarding default nor rate of return (p-values are 0.994, column 1, and 0.163, column 2). To further examine the randomization process based on borrower ex-ante characteristics, we replicate the analysis with the dependent variables as *Annual Income, Debt-to-Income Ratio,* and *Funding Amount*. The results are shown in Table 3.3, columns 3, 4, and 5 respectively. We find that the platform is randomizing with respect to these ex-ante variables as well (associated p-values are 0.997, 0.194, and 0.298 respectively).

3.6.2 Expertise of complementors in P2P Lending

Before investigating our main hypotheses, we provide an analysis regarding the institutional investors' capacity to utilize their expertise to select superior loans. We estimate the following specification:

$$Outcome_i = \beta_0 + \beta_1 * Loan\ Rejected\ by\ Institutional\ Investors_i + \gamma_i + \epsilon_i$$

where *Loan Rejected by Institutional Investors* has a value if 1 if the loan was initially assigned to the priority institutional investor market but was rejected by the institutional investors and recycled to the retail investors. Unlike our main specification, the sample we use in this analysis is from the loans that were initially assigned to the priority institutional loan market (including both loans rejected and funded by the institutional investors). There are 445,376 loans in this sample.

We report the results in Table 3.4. Column 1 and 2 show the results for *Default* and *Rate of Return* respectively. The results indicate that, on average, after controlling for the loan grade and origination year, loans rejected by the institutional investors have higher default rates and lower rates of return. The results provide empirical support that the loans rejected by institutional borrowers perform worse than the loans selected by the institutional investors.

3.6.3 Main Results

Moving to testing of hypothesis 1, we examine whether platform assigned level of privileged access has an impact on the conditional expectation of *Default* and *Rate of Return*. We utilize the following specification:

$$Outcome_{i\_t} = \beta_0 + \beta_1 * Proportion\ Priority\ Institutional\ Loans_t + \beta_2 * X_i + \gamma_{i\_t} + \epsilon_{i\_t}$$

where $Outcome_{i\_t}$ is the *Default* or *Rate of Return* for loan *i* which was originated in week *t*. The level of analysis is the individual loan (*i*). The sample used is the *Full Retail Market Sample*, which includes all loans that are made available to the retail investors (whether initially assigned or recycled from the priority institutional loans market). The *Proportion Priority Institutional Loans*$_t$ is the proportion of loans in the retail market, during week *t*, that were originally assigned to the priority institutional loans market. X_i refers to the set of explanatory borrower and loan specific characteristics outlined previously in Table 3.1. γ_{i_t} refers to the grade by year fixed effect corresponding to loan *i* and the year corresponding to week *t*.

Columns 1 and 2 of Table 3.5 show the results of this specification. Column 1 indicates that a 10% increase in the level of priority access increases the conditional probability of default in the retail loans market by approximately 0.16%. Given that the mean level of privileged access in our sample is 57%, this corresponds to an increase in default rate of approximately 0.91%. Column 2 indicates that the equivalent 57% priority access level leads to a decrease in return of approximately 0.051 percentage points. Compared to the average default rate on the platform, which is 14.65%, these changes represent a 6.2% change relative to the average. Similarly, the changes in rate of return represent an 8% increase compared to the average rate of return (6.4%). These results provide support for Hypothesis 1.

In hypothesis 2a, we posit that cream skimming by the experts has a negative spillover effect on the payoff of the crowd in online P2P lending platforms. To test this hypothesis, we utilize the following specification:

$$Outcome_i = \beta_0 + \beta_1 * Loan\ Rejected\ by\ Institutional\ Investors_i + \beta_2 * X_i + \gamma_i + \epsilon_i$$

Similar to the analysis in hypothesis 1, this specification is estimated at the individual loan level. As previously discussed, *Outcome$_i$* refers to either *Default* or *Rate of Return. Loan Rejected by Institutional Investors$_i$* is an indicator variable which has a value of 1 if the loan was originally assigned to the priority institutional loans market and rejected by the institutional investors. X_i refers to the previously described explanatory borrower and loan level characteristics, while γ_i refers to the grade by year dummy fixed effects.

If the loans rejected by the institutional investors negatively impact the retail market then we would observe a deterioration in the conditional expectation of the retail loan outcomes. Columns 3 and 4 of Table 3.5 present the results of this specification. The results in column 3 indicate that, on average, loans that are recycled from the institutional investor pool have a default rate 0.9 percentage points higher than the default rate of the loans that were initially assigned to the retail loan market by the platform. Moreover, column 4 shows that the loans rejected by the institutional investors have a rate of return that is 0.3 percentage points lower. Compared to the mean levels for default (rate of return), these changes represent an increase (decrease) of approximately 6.14% (4.7%). These results provide support for hypothesis 2a.

In hypothesis 2b, we postulate that an increase in exposure of the crowd to the residuals of cream skimming by the experts leads to higher same-side negative spillover effects in online P2P lending. We can measure the level of cream skimming as the proportion of the retail loan market in each week *t* that is comprised of loans that the institutional investors rejected. Specifically, we utilize the following specification:

$$Outcome_{i\_t} = \beta_0 + \beta_1 * Proportion\ Rejected\ Institutional\ Loans_t + \beta_2 * X_i + \gamma_{i\_t} + \epsilon_{i\_t}$$

This is analogous to the previous specifications except that the variable of interest is the *Proportion Rejected Institutional Loans* which corresponds the proportion of the retail market, in week *t*, that is comprised of loans rejected by institutional investors. This analysis is estimated at the individual loan level (*i*).

Columns 5 and 6 of Table 3.5 show the results of this specification. Column 5 shows that, for a 10% increase in the proportion of rejected loans that comprise the retail market, the estimated increase in default rate is 0.37 percentage point. Column 6 provides the analogous result for rate of return and estimates that a 10% increase in the proportion of rejected loans decreases the retail market rate of return by approximately 0.15 percentage points These results provide support for hypothesis 2b.

3.6.4 Supplementary Analyses

We conducted numerous analyses to examine the robustness of our results and broaden our understanding of the findings and mechanisms.

Falsification Test Using Institutional Loans Market.

One critical assumption of our specifications is that we do not have unobserved variables that simultaneously impact level of priority access and retail market outcomes. To examine the robustness of our findings, we replicate the analysis used to test hypothesis 1, except we use the sample of all loans that were initially assigned to the institutional investors. Given the random assignment of loans within grade and the controls we have included in our specification, we should not observe an association between the level of priority access (*Proportion Priority Institutional Loans)* and the outcomes in the loans randomly assigned to the institutional investors. Columns 1 and 2 of Table 3.6 present the results and, reassuringly, we do not find a relationship between the level of priority access and the outcomes in the sample comprised of loans randomly assigned to the institutional market (p-values are 0.473, column 1, and 0.166 column 2).

Evaluating Funding Time.

To delve further into the implications of changes in the level of priority access on the retail loans market, we examine the funding time for loans. Providing institutional investors priority access naturally reduces the size of the retail loans market, even though the overall pool is randomized. Therefore, the time to fund observed in the retail market can provide us insights into the impact of the level of priority access assigned by the platform.

We obtain the time difference between each individual loan request and the time it was funded for each loan in our sample starting in October 2014.[14] Before discussing the role of funding time as it relates to priority access, we examine whether funding time is related to loan outcomes in general. Specifically, we use the following specification:

$$Outcome_i = \beta_0 + \beta_1 * \log(Funding\ Time)_i + \beta_2 * X_i + \gamma_i + \epsilon_i$$

where $\log(Funding\ Time)_i$ is the log of the hours between the time when the loan is requested by the borrower and the time when the borrower receives their loan. X_i is the set of borrower and loan level explanatory variables and γ_i is the grade by year dummy that corresponds to loan (i). This approximates the time it took the loan to obtain funding. We focus on the *Full Retail Market Sample* as it is the focus of our study. The results are presented in columns 3 and 4 of Table 3.6. They suggest that loans that take longer to fund perform worse, on average, than loans that fund quickly.

Since funding time is relevant to loan outcomes, it is important to study how changes in platform assigned priority access impact the funding time. To evaluate this possibility, use the same specification used to test hypothesis 1, except we use $log(Funding\ Time)_{i\_t}$ as the dependent variable. The results are presented in column 5 of Table 3.6. The results indicate that a 10% increase in the level of privileged access results in 3.5% reduction in funding time in the retail market. Given the average level of privileged access is 57%, this indicates a decrease of approximately 20% in the funding time in the retail market when there is an average level of priority access prescribed by the platform. This provides evidence that, as the platform increases the level of priority access, the retail investors are forced to invest in the loans at a faster pace.

Given that the level of priority access decreases time to fund and time to fund impacts loan outcomes, it is natural to subsequently examine how funding time moderates the relationship between priority access and retail market outcomes (H1). To examine this we re-estimate the

[14] This information is unavailable before this period. Prior to this period, LC does not provide detail on the listing date and funding date of each loan.

specification used for hypothesis 1, except we include $log(Funding\ Time)_{i\_t}$ as a moderator for the level of priority access. We present the results in columns 6 and 7 of Table 3.6. The results indicate that, as the level of priority access increases, the loans that have a longer funding time are performing even worse.

Evaluating the Number of Investors

To understand further about how the priority access changes the number of investors needed to close the loans, we use number of investors in the log form as a dependent variable. We show in Table 3.7 that both intensity of ex ante priority access and cream skimming by institutional complementors, increases the numbers of retail investors needed to fund a loan. We posit that it is a form of risk diversification by the retail investors reducing the commitment to individual loans when they are exposed to the intensity of competition.

Heterogeneity Using Simulated Investors.

Thus far, we have identified that a how, LC reallocating the rejected whole loans to the retail investor pool, has a detrimental impact on the subset of participants with less expertise. However, to understand further the outcomes of this design choice on the retail investors, it is important to examine the heterogeneity of this result for different types of investors. Specifically, we ask the question: What are the characteristics of the investors that will be most affected by this design choice? To answer this question, we estimate the outcomes of five simulated investors during the period of our analysis.

We establish five investing strategies which range in terms of their associated risk profiles. Each strategy allocates a portion of loans from each loan grade (A, B, C, and D) to be invested in. Table 3.8 outlines the grade proportion allocations for each investment strategy. We examine the potential returns for each investment strategy using the following simulation approach:

1. For a specific month/year combination, starting with October 2014, we create a subset of all loans that originated in this period. We consider these loans as an investor's pool of possible investments. We remove all Purchased Whole Loans. As a result, we are left with loans that eventually entered the retail market and were funded fractionally. Note that these include both Retail Loans and Rejected Whole Loans.

2. The investor is allocated \$10,000 to invest. The \$10,000 is split into 100 shares of \$100 each, where each share is a fractional investment in a loan. The 100 loans are randomly chosen based upon the proportions outlined in Table 3.4. Using the rate of return calculations previously outlined for each loan, we determine the rate of return of the investor's \$10,000 investment.
3. We replicate Step 2 1,000 times and identify the average return for the investment across the 1,000 simulations. This is the investor's average return for a specific month/year combination.
4. We replicate steps 1-4 for all the month/year combinations (October 2014 - December 2015).

In the process outlined above the investor's pool of potential loans included Retail Loans and Rejected Whole Loans. We refer to this as Simulation 1. To examine the impact of the Rejected Whole Loans on the returns for each investment strategy, we replicate the steps outlined except we remove the Rejected Whole Loans from the pool of possible loans. We refer to this as Simulation 2. Figure 3.5 compares the returns for each scenario (including and not including Rejected Whole Loans) for each investment strategy. The shaded area in each graph indicates periods where the proportion of loans initially assigned to the whole loan pool is greater than 50%. Figure 3.5 clearly indicates that investment strategy 1 (from Table 3.8) is most impacted by the mechanism design choice to reallocate the Rejected Whole Loans to the fractional loan market. This effect is clearly evident in the periods where LC assigns more than 50% of loans to the whole loans market (the shaded region). The impact is less pronounced as the investment strategies become less dependent on loans with higher grade borrowers.

3.7 Discussion

3.7.1 Research Contributions

This study makes several research contributions. First, our study contributes to research on two-sided platforms by discussing the governance and internal regulatory role played by the platform owners (Boudreau & Hagiu, 2009; Evans & Schmalensee, 2013). Prior research has paid extensive attention to platform market access and emphasized platform owners ability to grant, deny, and remove access to the platform as a key dimension of platform governance (Boudreau, 2017; Parker & Van Alstyne, 2018). In a similar vein, we discuss how granting priority access to a subset of complementors affects the creation and capturing of value in the platform markets. We contribute to the literature by showing that, *within* the same side of the market, negative spillover effects and lemon market challenges arise when a governance mechanism such as priority access is deployed by the platform to possibly "grow the pie" and remove the potential bottlenecks in serving the demand-side of the market. We further elucidate the mechanisms that drive the results from priority access—the level of priority access granted to institutional players and cream skimming effects resulting from sending rejected priority institutional loans to the retail market, and how it can affect the value proposition for the relevant complementors. Our findings highlight how conflict of interest between the platform (granting priority access to institutional investors) and some set of complementors (retail investors having level-playing access to loan borrowers), can create challenging contingencies to the design of economic incentives in the platform markets.

Second, our study adds to a growing body of literature highlighting the importance of understanding the cohabitation of experts and crowd complementors in the two-sided platform markets (Boudreau, 2018; Greenstein & Zhu, 2018; Mollick & Nanda, 2016; Vallee & Zeng, 2019). We move beyond much of the research that discusses the conditions under which experts or crowds perform better, to the role that platform governance plays in creating contingencies on their performance, even after pricing for their expertise. We complement the current research by providing evidence for the interaction effect of platform design/governance mechanisms and the level of expertise of the complementors. Moreover, we unpack the challenges in designing and governing the platform market players with varying levels of expertise by linking such differences to one such governance mechanism, priority access, within broader platform governance framework.

Third, we coalesce the core corporate strategy literature focusing on governance mechanisms with the emerging platform markets (Adner *et al.*, 2018; Barach *et al.*, 2019; Chu & Wu, 2019). Platform market scope and boundaries are determined by the type and amount of complementors participating on both sides of the market (Boudreau, 2017; McIntyre & Srinivasan, 2017). Unlike traditional firms, platform owners don't have hierarchical (within firm) or strict contractual (interfirm) relationships with the complementors; instead they attempt to align for a value proposition together with the platform owner (Adner, 2017). By zooming in platform market access, and priority access in particular, we provide a useful addition to conventional corporate strategy, and show the emergence of possibly new governance mechanisms. This adds to the recent corporate strategy discussion on the scale, scope and organization of platform, and how well-known corporate instruments such as decision rights, ownership structures and process control methods are helpful for managers in achieving desired outcomes (Adner *et al.*, 2018; Chu & Wu, 2019; McIntyre & Srinivasan, 2017).

3.7.2 Practical Implications

Our research has several implications for practice. First, as Roth (2018) notes, the ability to design the marketplaces call for a "microeconomic engineering" thinking from the managers and strategists. By empirically evaluating how design choices such as priority access can create downstream effects on the complementors, we invite platform designers to delve further into both the first order—how the governance mechanisms such as priority access is deployed—and second order—how the spillover effects can emerge. According to our research, a fuller understanding of the incentive structures and the *changes* design and deployment of governance mechanisms can bring to them is critical to obtaining a holistic picture about the platform.

Second, our findings provide insights regarding the antitrust issues potentially faced by the platform. The two-sided nature of the platforms require changes to the existing antitrust apparatus and a "refrain from mechanically applying standard antitrust ideas where they do not belong" (Tirole, 2015). One particular challenge is that platforms often create substantial value for the complementors, while often dictating how the value gets appropriated among them (Evans & Schmalensee, 2013; Stigler Committee on Digital Platforms, 2019; Zhu & Liu, 2018). However, the analysis on the suitable ways to use the regulatory apparatus remains primitive. Stigler Committee on Digital Platforms (2019) lamented this limitation as follows: "the proposals were reactions to the perceived threat posed by digital platforms, with little to no analysis of the underlying root problems, let alone a link between market failures and remedies". In our paper, we show how platform owners' governance mechanisms can create negative spillover effects on the crowd market players even as they proclaim them to be the "heart" of the platform. The set of nuanced implications indicated here open up some potential pathways to thinking further about the antitrust issues coming from value creation and appropriation in the platform markets.

3.7.3 Limitations and Future Research

Several limitations in this study point to future research directions. First, although we found that granting privilege access to institutional investors bring negative spillover effects on the same-side retail investors, we haven't been able to identify different types of institutional investors in the platform (such as investment banks, pension funds, commercial trusts, endowment funds, hedge funds, and private equity investors) due to data limitations. Prior research in financial markets show that there are performance heterogeneities among the type of institutional investors. Future work, therefore, should examine them in this context.

Second, as noted in the paper, platform owner in our context, Lending Club, resorts to providing priority access to institutional investors with the purpose of removing the demand-side bottlenecks in the pursuit of expanding the platform. This design approach is followed LC's main rival Prosper Marketplace in the United States as well. However, this also leads to a salient question on the counterfactuals: what are the alternative solutions to bringing credible institutional investors onboard without granting them priority access? Although, we show the implications of such policy, scholars can go beyond merely observing them to proposing viable alternative solutions that we suspect would be of interest to wider audience including strategists, market designers, economists and legal experts.

Third, while our context of online marketplace lending is well suited to study the interplay between platform governance and the level of expertise, future work can extend this research to other platforms. For instance, we draw from the finance literature to show, *ceteris paribus*, how expert institutional investors perform better than the non-expert crowd investors, which in turn helps us to build the theory for the priority access. However, this effect may not be clear in other marketplaces. What are the contingency effects on the platforms where the collective intelligence of the crowds is better than the expert evaluation? We leave those extensions for future research.

3.8 Conclusion

Platform owners often align complementors with varied incentives in two-sides of the market towards orchestrating a specific value proposition. Emerging scholarship on platform governance suggests that platforms involve in market designing activities including the design of economic incentives to shape specific outcomes in the marketplace. Our analysis focuses on platform access control, specifically in the form of granting priority access to a subset of supply side complementors, and shows that the interplay between platform governance and the level of expertise can create important variations in complementor performance. Using online P2P lending industry, we demonstrate how platform market design choices related to lender side expansion, through giving priority access to institutional lenders, can have negative spillover effects on the retail lenders. We specifically discuss two mechanisms pertaining to priority access, the level of priority access and cream skimming in priority market, to explain further about the effects. As platform-based organizations proliferate in the digital economy, we hope our study will encourage further interest in platform market design, expertise of the market players, and corporate strategy.

3.9 References

Adner R. 2017. Ecosystem as Structure: An Actionable Construct for Strategy. *Journal of Management* **43**(1): 39-58.

Adner R, Puranam P, Zhu F. 2018. Special Issue of Strategy Science: Strategy in the Digital Era. *Strategy Science*.

Anderson Jr EG, Parker GG, Tan B. 2014. Platform Performance Investment in the Presence of Network Externalities. *Information Systems Research* **25**(1): 152-172.

Baldwin CY, Woodard CJ. 2009. The Architecture of Platforms: A Unified View. *Platforms, markets and innovation* **32**.

Barach MA, Kaul A, Leung MD, Lu S. 2019. Strategic Redundancy in the Use of Big Data: Evidence from a Two-Sided Labor Market. *Strategy Science* **4**(4): 298-322.

Biais B, Foucault T, Moinas S. 2015. Equilibrium Fast Trading. *Journal of Financial Economics* **116**(2): 292-313.

Bolton P, Santos T, Scheinkman JA. 2016. Cream-Skimming in Financial Markets. *The Journal of Finance* **71**(2): 709-736.

Boudreau K. 2010. Open Platform Strategies and Innovation: Granting Access Vs. Devolving Control. *Management Science* **56**(10): 1849-1872.

Boudreau KJ. 2012. Let a Thousand Flowers Bloom? An Early Look at Large Numbers of Software App Developers and Patterns of Innovation. *Organization Science* **23**(5): 1409-1427.

Boudreau KJ. 2017. Platform Boundary Choices & Governance: Opening-up While Still Coordinating and Orchestrating. *Entrepreneurship, Innovation, and Platforms. Advances in Strategic Management* **37**: 227-297.

Boudreau KJ. 2018. Amateurs Crowds & Professional Entrepreneurs as Platform Complementors, National Bureau of Economic Research.

Boudreau KJ, Hagiu A. 2009. Platform Rules: Multi-Sided Platforms as Regulators. *Platforms, Markets and Innovation* **1**: 163-191.

Butler AW, Cornaggia J, Gurun UG. 2017. Do Local Capital Market Conditions Affect Consumers' Borrowing Decisions? *Management Science* **63**(12): 4175-4187.

Casadesus-Masanell R, Hałaburda H. 2014. When Does a Platform Create Value by Limiting Choice? *Journal of Economics & Management Strategy* **23**(2): 259-293.

Chu LY, Wu B. 2019. Designing Online Platforms for Offline Services: The Bigger, the Better? . *Working Paper, University of Southern California and University of Michigan*.

Cohen JE. 2017. Law for the Platform Economy. *UC Davis Law Review* **51**: 133.

Constantinides P, Henfridsson O, Parker GG. 2018. Introduction-Platforms and Infrastructures in the Digital Age. *Information Systems Research* **29**(2): 381-400.

Cumming D, Meoli M, Vismara S. 2019. Investors' Choices between Cash and Voting Rights: Evidence from Dual-Class Equity Crowdfunding. *Research Policy* **48**(8): 103740.

Dushnitsky G, Piva E, Rossi-Lamastra C. 2020. Mix and Match? Exploring the Strategy Mix and Subsequent Performance of Transaction Platforms. *Strategic Management Journal* **Forthcoming**.

Eisenmann T, Parker G, Van Alstyne MW. 2006. Strategies for Two-Sided Markets. *Harvard Business Review* **84**(10): 92.

Evans DS. 2012. Governing Bad Behavior by Users of Multi-Sided Platforms. *Berkeley Technology Law Journal* **27**: 1201.

Evans DS, Schmalensee R. 2013. The Antitrust Analysis of Multi-Sided Platform Businesses, National Bureau of Economic Research.

Fishman MJ, Parker JA. 2015. Valuation, Adverse Selection, and Market Collapses. *The Review of Financial Studies* **28**(9): 2575-2607.

Glode V, Green RC, Lowery R. 2012. Financial Expertise as an Arms Race. *The Journal of Finance* **67**(5): 1723-1759.

Goldfarb A, Tucker C. 2019. Digital Economics. *Journal of Economic Literature* **57**(1): 3-43.

Greenstein S, Zhu F. 2018. Do Experts or Crowd-Based Models Produce More Bias? Evidence from Encyclopædia Britannica and Wikipedia. *MIS Quarterly*.

Halaburda H, Jan Piskorski M, Yıldırım P. 2018. Competing by Restricting Choice: The Case of Matching Platforms. *Management Science* **64**(8): 3574-3594.

Helfat CE, Raubitschek RS. 2018. Dynamic and Integrative Capabilities for Profiting from Innovation in Digital Platform-Based Ecosystems. *Research Policy* **47**(8): 1391-1399.

Jeppesen LB, Frederiksen L. 2006. Why Do Users Contribute to Firm-Hosted User Communities? The Case of Computer-Controlled Music Instruments. *Organization Science* **17**(1): 45-63.

Kahn AE. 1988. *The Economics of Regulation: Principles and Institutions*. MIT press.

Kim K, Viswanathan S. 2018. The Experts in the Crowd: The Role of Experienced Investors in a Crowdfunding Market. *MIS Quarterly*.

Laffont J-J, Tirole J. 1993. *A Theory of Incentives in Procurement and Regulation*. MIT press.

Laffont JJ, Tirole J. 1990. Optimal Bypass and Cream Skimming. *American Economic Review* **80**(5): 1042-1061.

Lending Club. 2012. How Lendingclub Balances Different Investors on Its Platform. https://help.lendingclub.com/hc/en-us/articles/115009000328-How-LendingClub-balances-different-investors-on-its-platform (2019-04-27 2019).

McIntyre DP, Srinivasan A. 2017. Networks, Platforms, and Strategy: Emerging Views and Next Steps. *Strategic Management Journal* **38**(1): 141-160.

Mollick E, Nanda R. 2016. Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts. *Management Science* **62**(6): 1533-1553.

Morse A. 2015. Peer-to-Peer Crowdfunding: Information and the Potential for Disruption in Consumer Lending. *Annual Review of Financial Economics* **7**: 463-482.

Parker G, Van Alstyne M. 2018. Innovation, Openness, and Platform Control. *Management Science* **64**(7): 3015-3032.

Poetz MK, Schreier M. 2012. The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas? *Journal of Product Innovation Management* **29**(2): 245-256.

PWC. 2015. Peer Pressure: How Peer-to-Peer Lending Platforms Are Transforming the Consumer Lending Industry. https://www.pwc.com/us/en/industries/banking-capital-markets/consumer-finance/library/peer-to-peer-lending.html (2019-02-19 2019).

Rochet J-C, Tirole J. 2003. Platform Competition in Two-Sided Markets. *Journal of the European Economic Association* **1**(4): 990-1029.

Rochet JC, Tirole J. 2006. Two-Sided Markets: A Progress Report. *The RAND Journal of Economics* **37**(3): 645-667.

Romanyuk G, Smolin A. 2019. Cream Skimming and Information Design in Matching Markets. *American Economic Journal: Microeconomics* **11**(2): 250-276.

Roth AE. 2018. Marketplaces, Markets, and Market Design. *American Economic Review* **108**(7): 1609-1658.

Rysman M. 2009. The Economics of Two-Sided Markets. *The Journal of Economic Perspectives* **23**(3): 125.

Shleifer A, Summers LH. 1990. The Noise Trader Approach to Finance. *Journal of Economic Perspectives* **4**(2): 19-33.

Stigler Committee on Digital Platforms. 2019. University of Chicago: Stigler Center for the Study of the Economy and the State.

Tech Crunch. 2018. China's Didi Chuxing Is Close to Launching a Taxi-Booking Service in Japan. https://techcrunch.com/2018/07/18/didi-taxi-japan/.

Teece DJ. 2018. Profiting from Innovation in the Digital Economy: Enabling Technologies, Standards, and Licensing Models in the Wireless World. *Research Policy* **47**(8): 1367-1387.

Tirole J. 2015. Market Failures and Public Policy. *American Economic Review* **105**(6): 1665-1682.

Tiwana A. 2013. *Platform Ecosystems: Aligning Architecture, Governance, and Strategy*. Newnes.

Tiwana A, Konsynski B, Bush AA. 2010. Research Commentary—Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics. *Information Systems Research* **21**(4): 675-687.

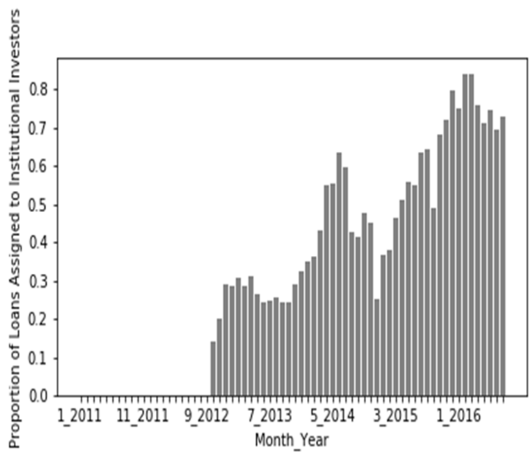
Trans Union. 2017. Fact or Fiction: Are Fintechs Different Than Other Lenders? https://www.transunion.com/blog/fact-or-fiction-are-fintechs-different-than-other-lenders (2019-04-27 2019).

Vallee B, Zeng Y. 2019. Marketplace Lending: A New Banking Paradigm? *The Review of Financial Studies* **32**(5): 1939-1982.

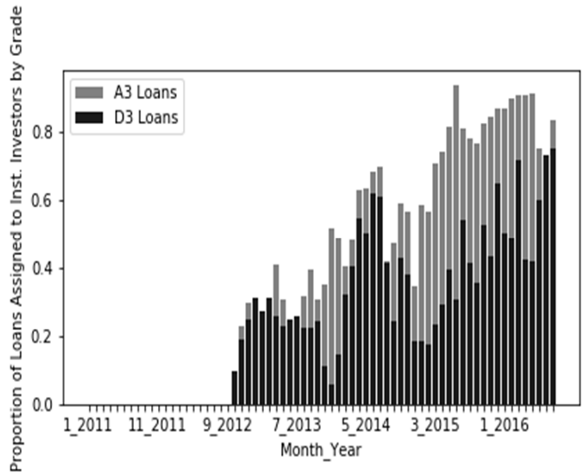
Vulkan N, Roth AE, Neeman Z. 2013. *The Handbook of Market Design*. Oxford University Press: Oxford.

Yoo Y, Henfridsson O, Lyytinen K. 2010. Research Commentary—the New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research* **21**(4): 724-735.

Zhu F, Liu Q. 2018. Competing with Complementors: An Empirical Look at Amazon. Com. *Strategic Management Journal* **39**(10): 2618-2642.



(Panel I)



(Panel II)

Figure 3.1 Monthly Institutional Loan Assignment Distribution

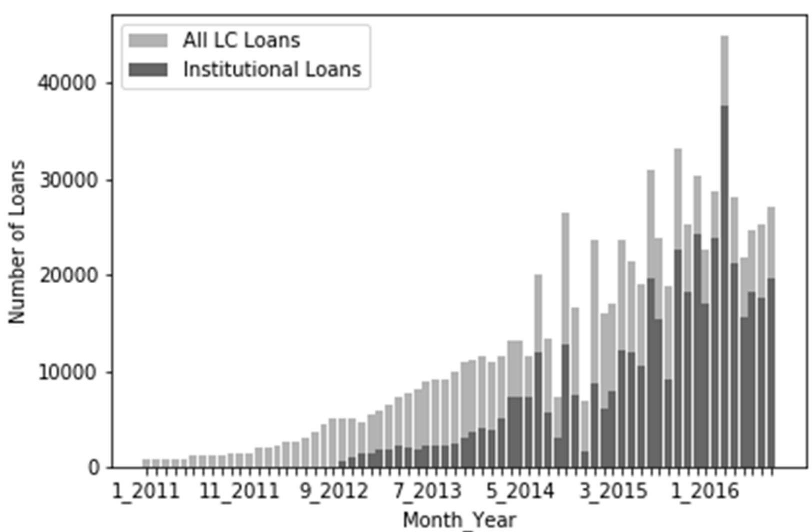


Figure 3.2 Lending Club Number of Loans Overall and Number of Loans Assigned to Institutional Investors

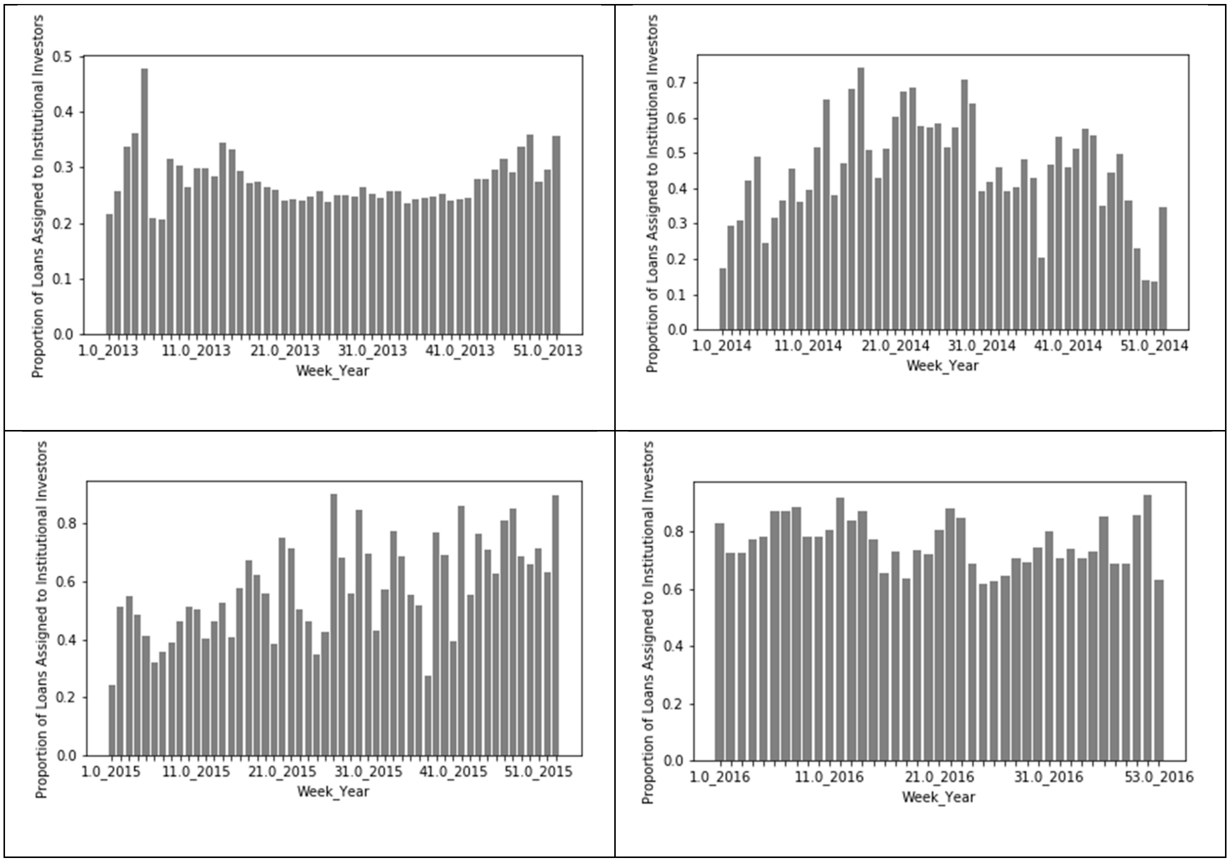


Figure 3.3 Weekly Institutional Loan Distribution by Year

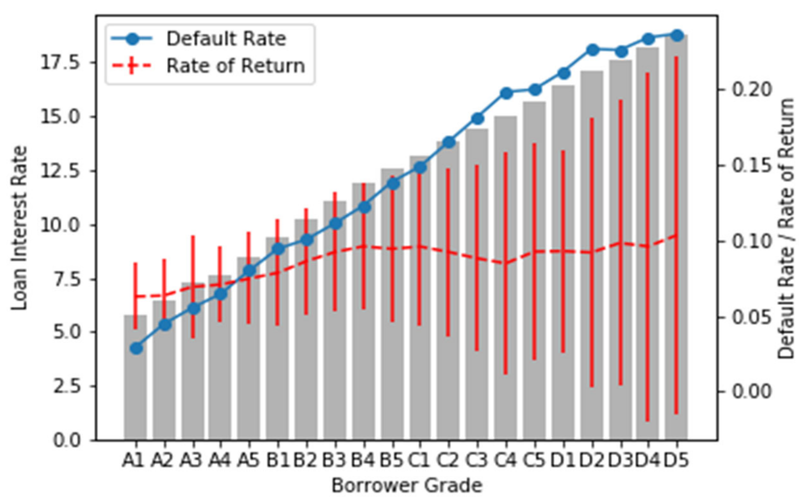


Figure 3.4 Interest Rate, Default, and Loan Rate of Return by Borrower Grade

Note: The bar graph in this figure represents the average loan interest rate for each borrower grade. The dashed line indicates the rate of return with the associated error bars representing the variance. The solid line represents the average default rate of each borrower grade.

Table 3.1Variable Description

	Variable	Description
Dependent Variables	*Default*	1: if the borrower does not default on loan 0: if the borrower defaults on loan
	Rate of Return	$\frac{Sum\ of\ Payments\ made\ by\ Borrower}{Funded\ Amount}$
Explanatory Variables	*Proportion Priority Institutional Loans*	For each week, the proportion of all LC listings that are assigned to the institutional investor market.
	Loan Rejected by Institutional Investors (1/0)	1: if the loan is initially assigned to the institutional investor market but subsequently rejected by the institutional investors. These loans are reassigned to the retail market. 0: all other loans, including loans initially assigned to the retail market and loans initially assigned to and funded by the institutional investors.
	Proportion Loans Rejected by Institutional Investors	For each week, the proportion of loans funded by the retail investors that were originally assigned to the priority institutional investor market.
Control Variables	*Funding Amount*	The log of the dollar amount of funding obtained by the borrower.
	Debt-to-Income Ratio	The ratio of the borrower's debt to their income prior to obtaining their LC loan.
	Inquiries Last 6 Months	The number of credit inquiries made by the borrower in the months prior to their LC loan request.
	Annual Income	The log of the annual income of the borrower (in $1,000s).
	Revolving Balance	The log of the revolving credit balance owed by the borrower prior to requesting a LC loan (in $1,000s).
	Interest Rate	The interest rate assigned by LC for the borrower's loan.
	Homeownership Status	1: Borrower rents their home. 0: Borrower owns home.
	Current Credit Lines	The log of the number of current credit lines the borrower has before obtaining their LC loan.

Table 3.2 Descriptive Statistics and Correlations

	Variables	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11	12
1	*Default*	0.16	0.36												
2	*Rate of Return*	0.07	0.26	**-0.85**											
3	*Proportion Priority Institutional Loans*	0.49	0.21	**0.06**	**-0.07**										
4	*Loan Rejected by Institutional Investors (1/0)*	0.11	0.31	0.00	**-0.02**	**0.29**									
5	*Proportion Loans Rejected by Institutional Investors*	0.11	0.15	**-0.01**	**0.02**	**-0.19**	**0.15**								
6	*Funding Amount*	12.67	8.02	**-0.01**	**-0.01**	**0.02**	**0.03**	**0.00**							
7	*Debt-to-Income Ratio (%)*	18.44	8.97	**0.09**	**-0.04**	**0.10**	**0.07**	**0.00**	**-0.01**						
8	*Inquiries Last 6 Months*	0.72	0.99	**0.07**	**-0.03**	**-0.02**	**0.02**	**0.03**	**-0.03**	0.00					
9	*Annual Income*	71.03	62.91	**-0.05**	**0.02**	0.00	**-0.02**	**-0.01**	**0.36**	**-0.19**	**0.04**				
10	*Revolving Balance*	15.56	23.01	**-0.04**	**0.02**	**0.00**	**0.03**	**0.01**	**0.33**	**0.11**	**-0.01**	**0.30**			
11	*Interest Rate*	12.75	4.11	**0.20**	**-0.02**	**-0.03**	**-0.13**	**0.02**	**-0.09**	**0.16**	**0.25**	**-0.13**	**-0.10**		
12	*Homeownership Status*	0.44	0.50	**0.07**	**-0.03**	**0.01**	**-0.01**	0.00	**-0.16**	**-0.01**	**-0.04**	**-0.14**	**-0.15**	**0.13**	
13	*Current Credit Lines*	24.49	11.82	**-0.02**	**-0.02**	**-0.01**	**0.03**	**-0.01**	**0.21**	**0.20**	**0.16**	**0.19**	**0.18**	**-0.12**	**-0.21**

N = 378,487. All bold values are significant at the $p<0.05$ level, two-tailed test.

Table 3.3 Analysis of LC Randomization of Loan Assignment

	(1)	(2)	(3)	(4)	(5)
DV:	*Default*	*Rate of Return*	*Annual Income*	*Debt-to-Income Ratio*	*Funding Amount*
Priority Institutional Loan (1/0)	-0.000 (0.001)	-0.001 (0.001)	0.007 (0.170)	0.028 (0.021)	0.020 (0.020)
Grade by Year FE	Yes	Yes	Yes	Yes	Yes
Constant	0.040 (0.016)	0.056 (0.012)	84.882 (3.239)	12.741 (0.410)	11.178 (0.376)
Num. obs.	783,999	783,999	783,999	783,999	783,999

Standard errors are reported in parentheses.

***Notes*:** This table reports the results of an examination of LC's randomization process in allocating loans between the retail market and the priority institutional loans market. The results indicate that, on average, there is no difference in loans assignment based on *Default, Rate of Return, Annual Income, Debt-to-Income Ratio,* and *Funding Amount*.

Table 3.4 Analysis of LC Institutional Investor Performance

DV:	(1) *Default*	(2) *Rate of Return*
Rejected Institutional Investor Loan (1/0)	0.023 (0.002)	-0.011 (0.001)
Grade by Year FE	Yes	Yes
Constant	0.008 (0.037)	0.070 (0.026)
Num. obs.	445,376	445,376

Standard errors are reported in parenthesis.

***Notes*:** This table provides the performance of the institutional investors. Specifically, the results compare the differences between the outcomes of the loans funded by the institutional investors and those rejected by the institutional investors. The results indicate that the loans rejected by institutional investors defaulted more often and had a lower rate of return than the loans they funded.

Table 3.5 Main Results

	H1: Level of Priority Access		H2a. Cream Skimming		H2b. Level of Cream Skimming	
	(1)	(2)	(3)	(4)	(5)	(6)
DV:	*Default*	*Rate of Return*	*Default*	*Rate of Return*	*Default*	*Rate of Return*
Proportion Priority Institutional Loans	0.016 (0.004)	-0.009 (0.003)				
Loan Rejected by Institutional Investors (1/0)			0.009 (0.002)	-0.003 (0.001)		
Proportion Loans Rejected by Institutional Investors					0.037 (0.007)	-0.015 (0.005)
Funding Amount	0.032 (0.001)	-0.019 (0.001)	0.032 (0.001)	-0.019 (0.001)	0.032 (0.001)	-0.019 (0.001)
Debt to Income Ratio	0.001 (0.000)	-0.001 (0.000)	0.001 (0.000)	-0.001 (0.000)	0.001 (0.000)	-0.001 (0.000)
Inquiries Last 6 Months	0.011 (0.001)	-0.010 (0.000)	0.011 (0.001)	-0.010 (0.000)	0.011 (0.001)	-0.010 (0.000)
Annual Income	-0.038 (0.002)	0.018 (0.001)	-0.037 (0.002)	0.017 (0.001)	-0.038 (0.002)	0.018 (0.001)
Revolving Balance	-0.016 (0.002)	0.016 (0.001)	-0.016 (0.002)	0.016 (0.001)	-0.016 (0.002)	0.016 (0.001)
Interest Rate	-0.001 (0.002)	0.010 (0.001)	-0.001 (0.002)	0.011 (0.001)	-0.001 (0.002)	0.011 (0.001)
Homeownership Status	0.031 (0.001)	-0.018 (0.001)	0.031 (0.001)	-0.018 (0.001)	0.031 (0.001)	-0.018 (0.001)
Current Credit Lines	0.009 (0.001)	-0.016 (0.001)	0.008 (0.001)	-0.016 (0.001)	0.009 (0.001)	-0.016 (0.001)
Grade by Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.001 (0.020)	0.058 (0.015)	0.005 (0.020)	0.055 (0.015)	-0.026 (0.021)	0.070 (0.015)
Num. Obs.	378,487	378,487	378,487	378,487	378,487	378,487
R-Squared	0.057	0.019	0.057	0.019	0.057	0.019

Standard errors are reported in parenthesis.

***Notes*:** This table presents the main results examining the hypotheses proposed. Columns 1 and 2 examine hypothesis 1, columns 3 and 4 hypothesis 2a, and columns 5 and 6 hypothesis 2b.

Table 3.6 Supplementary Analysis – Funding Time

	Falsification Test		**Evaluating Funding Time**				
DV:	(1) *Default*	(2) *Rate of Return*	(3) *Default*	(4) *Rate of Return*	(5) *log(Funding Time)*	(6) *Default*	(7) *Rate of Return*
Proportion Priority Institutional Loans	0.003 (0.004)	-0.004 (0.003)			-0.359 (0.007)	-0.036 (0.047)	0.038 (0.031)
Funding Time			0.018 (0.002)	-0.007 (0.001)		0.012 (0.005)	-0.001 (0.004)
Proportion Priority Institutional Loans x log(Funding Time)						0.012 (0.009)	-0.010 (0.006)
Loan Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade by Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.052 (0.038)	0.092 (0.027)	-0.160 (0.022)	0.167 (0.015)	5.731 (0.028)	-0.151 (0.036)	0.150 (0.024)
Num. Obs.	445,338	445,338	223,461	223,461	223,461	223,461	223,461

Standard errors are reported in parenthesis.
***Notes*:** This table provides results from the supplementary analysis. Funding Time refers to the log of the time between when the borrower listed their loan and when they received their loan.

Table 3.7 Supplementary Analysis – Number of Investors

DV:	***Number of Investors (in Log)***	
Proportion Priority Institutional Loans	0.371 (0.008)	
Proportion Loans Rejected by Institutional Investors		0.185 (0.014)
Controls	Yes	Yes
Grade by Year FE	Yes	Yes
Num. Obs.	416,515	416,515

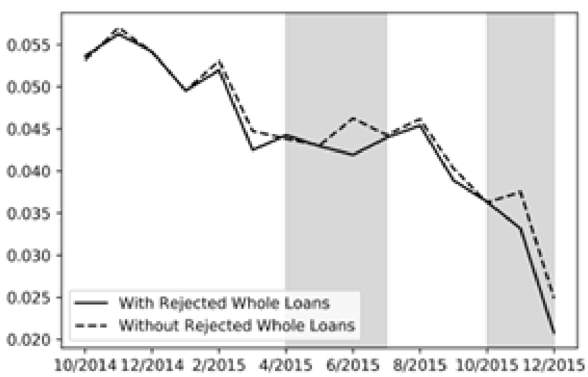
Standard errors are reported in parenthesis.

Notes: This table provides results from the supplementary analysis. Funding Time refers to the log of the time between when the borrower listed their loan and when they received their loan.

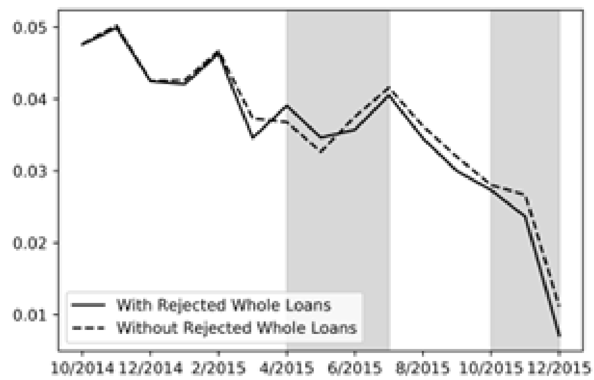
Table 3.8 Simulated Investor Rate of Return

Investor 1

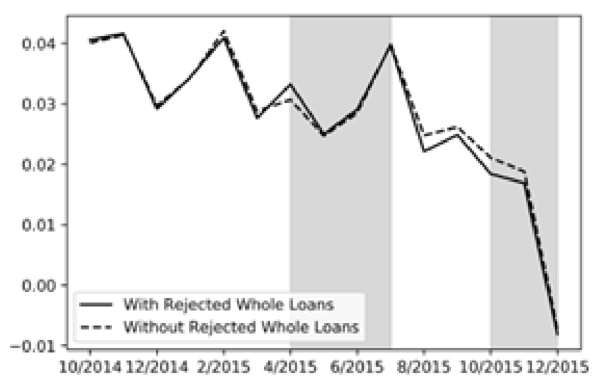
Investor 2



Investor 3



Investor 4



Investor 5