1. Introduction

During the COVID-19 pandemic, many people were kept far away from colleagues, co-authors, and co-founders, but their work continued via video conferencing, smart phones, and email—something that would not have been possible only decades ago. From the printing press to the Internet, technology has long helped conquer the tyranny of distance and time. Increasingly, this is true not only for sharing ideas but also for creating them. Innovation in science and technology is becoming an asynchronous conversation across time, geographies, organizations, and disciplines. Existing frameworks for how these innovations occur, however, are mired in language about cities, countries, companies, founders, and universities.

Scientific and technological progress is increasingly the result of a profoundly collaborative endeavor—split between innovators in the private and the public sectors, large companies and startups, universities and technology hubs, colleagues and strangers, and those near and far. The stories we tell about such innovation in both the academic literature and popular discourse do not capture these rich interactions well. Research in academia and management consulting too often ends up commodifying innovation—focusing on generic “firms” and modeling them as machine-like entities to be studied in terms of internal dynamics and key differences. Popular narratives in the press, even in Silicon Valley, are too eager to find a hero—focusing only on the *de novo*, singular brilliance of a technical founder, engineer, or small team of scientists.

An underappreciated and understudied part of innovation in science and technology is the contribution of spontaneous clusters of researchers, scientists, and thinkers, around the world, working on a common project, united by common purpose, not connected by organizational or other affiliation. Such rich and generative environments today have critical social dimensions that are hidden in plain sight. Indeed, the most important teams in the world of science and technology may already be teams of strangers. Much of the key work in innovation for technical fields is not coordinated by universities, governments, big tech, or any other formal institutions. This is the phenomenon we seek to better understand.

The special power of informal collaboration between researchers is an old idea, often traced to Bernard of Chartres in the twelfth century and known best in the words of Sir Isaac Newton: “If I have seen further, it is by standing on the shoulders of giants.” Such collaboration has expanded exponentially with the advent of the Internet, the digitalization of knowledge, and the globalization of research and technology. Modern evangelists of blockchain envision a decentralized Web3 version of the Internet rebuilt on such non-hierarchical networks and governed via shared rules. But until the last few decades, the technology did not exist to study such broad collaboration or understand it systematically. Recently, however, technologies developed to, for example, cluster similar web pages and analyze massive quantities of data, like bioinformatics of gene expression, allow us to anecdotally recognize the importance of informal col-

laboration and to systematically study the mechanisms that lead to success. Perhaps someday we can even learn how to encourage such interactions to accelerate innovation.

In some rare, special cases, the work of a group of loosely associated innovators, working collectively but not together, assumes a sort of joint coherency. The group begins to function as if it were a new emergent organizational form—capable of very productive collaboration. We refer to this phenomenon as a “knowledge community.” While it is not exclusive to innovation in science and technology, we think it is particularly prevalent there. Scientific or technological knowledge communities comprise an interorganizational large-scale network in which researchers work together by building on each other’s advances—whether in an approach to cancer research, a new microchip, or a COVID-19 vaccine. In the for-profit realm, these communities often form in times of urgency, technological breakthrough, or existential need—when sharing ideas quickly trumps keeping them proprietary, and good ideas are more important than firm borders.

These knowledge communities are an exemplar of the social process we are studying. Many studies suggest that researchers in such unstructured research clusters or communities tend to produce at much higher rates of innovation than either less cohesive researchers or more structured teams. But rarely has research taken such knowledge communities as central, not only to the success of specific innovations, but to the continued success of the processes that generate those innovations.

The implications of thinking about networks of innovators who share ideas instead of formal organizations (like companies, governments, universities, or nonprofits) are profound—indeed, we find that unusual organizing logics guide such informal research clusters. Existing anecdotal research suggests, to generalize very broadly, that organizations such as technology firms seek to combine diverse specialties and to coordinate deeply heterogeneous tasks—such as marketing, finance, management, and production—toward a set of goals. To generalize again, networks, such as knowledge communities, however, tend toward homogeneity by gathering together people with similar ideas where they can learn from and build upon each other’s work. How these two different patterns of influence and innovation overlay each other in dynamic tension in the real world, where both exist simultaneously, is particularly salient in arenas such as R&D and technology.

Examining the interplay between organizational boundaries and knowledge flow reveals interesting and counterintuitive lessons about identifying promising innovations as they emerge. How do successful knowledge creators, CEOs, and corporate R&D researchers engage in communal knowledge development and simultaneously maintain proprietary knowledge? Understanding this would inform the corporate and scientific use of patents, internal knowledge development, and more generally the way companies might identify and take advantage of innovation and change. Since we find that areas of innovation often originate in such decentralized intellectual communities, it may be advantageous for firms to identify them as they emerge in order to benefit from their good ideas, to monitor the risks and rewards of investing in a community or an idea, and to be better aware of how their scientists are interacting with knowledge communities. Indeed, this is precisely how the innovations that underpin Google originated.

THEORY AND DATA

Network theory over the past several decades has revealed the difference interpersonal network ties make, and how crucial they are to social structures. Seminal network theorists James Coleman and Ron Burt challenged traditional views about what determines outcomes—demographic attributes, such as class, age, and strength—with network-based ideas, such as network centrality, social capital, and structural holes (Burt, 2000). Network theorists argued that, in many cases, outcomes are determined not chiefly by what you know, but instead by whom you know and how you are connected to others. Later network researchers, such as Harrison White, Mark Granovetter, and more recently Duncan Watts, have built on these ideas by showing that the structure of relationships does matter, not only in a given situation but also on the larger scale of social structure—which they see as comprising interpersonal/internode relationships writ large (Burt, 1977; Granovetter, 1994; Jacobs & Watts, 2021; Watts, 1999; White, 2003).

We use data to identify and understand, at least on some level, these intense clusters of innovators that we call knowledge communities. The challenge, formally speaking, in researching this new informal model of collaboration, is to infer the social structure of knowledge from the content of interactions. This book looks at the use of shared knowledge, or shared rhetoric, as the unit of agglomeration for the social structure of knowledge. In our data context—citations and language in scientific and social science publications—content is more apparent and accessible. Therefore, rather than inferring content from connections, we infer connections from content. These connections, or ties, however, are not friendship ties, or even necessarily acquaintanceship ties, but exchanges of ideas (or proxies for them, at least). We show how ideas are tied together into knowledge communities and how their position in these communities has consequences for their impact.

The difficulty of finding and interpreting the content of interactions is increasingly reduced by technologies, such as email and inexpensive electronic storage, which have opened up new avenues for content analysis. One very promising area for exploration is the systematic compilation of text and citations in shared academic discourse, such as the sciences and social sciences, which is where we focus our efforts (Small, 1992, 1998, 2018b; Upham, Rosenkopf, & Ungar, 2010a, 2010b; Upham & Small, 2010). In other words, we use techniques in network analysis to study the content of interactions and the social structure implied by that content.

Scientific and academic discourse are of special interest because of particular attributes in this arena and the more public way in which ideas are exchanged. Three aspects of scientific articles are particularly useful: First, published articles in the sciences and social sciences describe the researcher's own positions. Second, published articles include citations through which the author acknowledges the influence of the knowledge and ideas of others. Third, these published articles can themselves be cited, becoming a part of this social discourse of ideas. The same is true for patent research, though secrecy and legal considerations make this context somewhat more complex.

The formal work of scientists and technologists—their papers, patents, and inventions—becomes their primary source of professional self-identification and in turn helps shape their intellectual community (Hargens, 2000b; McCain, 1986a; Parker & Corte, 2017; Schardosin et al., 2020). The printed page or, increasingly, the computer screen is where scientists interact. Before the Internet and long-distance telephony, scholars around the world rarely, if ever, met,

interacting via the printed page. Scientists are not, in other words, intellectually defined so much by with whom they have lunch as by with whom they agree intellectually and on whose innovations and work they build (Merton, 1965). Of course, the social structure in intellectual communities can be complemented by personal interactions at conferences and universities. Nevertheless, scientists are often more similar, in terms of discipline, knowledge and ideas, to someone with whom they agree intellectually and have never met than they are to the researcher across the hall (Collins, 1997). In other words, in technical domains, it is often their knowledge community, not just their organization, that shapes a person's innovations (Zhang et al., 2020).

SUMMARY

This book attempts to study the interaction between knowledge communities and organizational boundaries in different contexts, in order to better understand how they intersect in practice. We focus on three communities of innovators and examine, in each, different aspects of how knowledge is exchanged and how broad collaboration has changed old knowledge-sharing paradigms. In three chapters, we emphasize the content of interactions and their effect on performance by examining citations in academic and scientific papers. We also focus on the methodology of clustering in order to find intellectual communities formed on the basis of ideas and beliefs. Finally, we focus on the importance of considering method in evaluating the coherency and validity of knowledge communities (Upham, 2006; Upham, Rosenkopf, & Ungar, 2010a).

We study, in Chapter 2, the field of computer science over more than a decade. We ask how computer scientists organize themselves into cohesive knowledge communities and how these communities help their members to succeed or fail (Upham, Rosenkopf, & Ungar, 2010b). In this field, there is a good deal of direct involvement from the private sector, and many discoveries have led directly to entrepreneurial and business applications (Ma & Uzzi, 2018; Popescul et al., 2000). Computer science as an area of research is a “dual economy,” where incentives to publish and generate publicly useful information can coexist with the incentive to use information in a proprietary way to gain direct market returns (Gittelman & Kogut, 2003). That science and industry are more directly intertwined in this context yields an interesting angle for analysis.

As we step back to examine the way in which knowledge communities interact with each other, we begin to examine patterns and develop explanations for their differential performance—explanations that differ significantly from intuitions developed from looking at interactions between formal organizations. These knowledge communities are not led by a CEO or president; they do not have an official name or organizational structure. They exist, rather, as a reflection of the substantiveness of their members' knowledge-sharing behavior—the cohesiveness of their shared content. There is a tension between the way these communities use knowledge and the way they generate rhetoric—a tension that is irresolvable because these two aspects of knowledge communities are intertwined and yet pull in orthogonal directions.

Little work has looked at what drives the relative performance of knowledge communities and indirectly, therefore, the relative progress of science. Understanding knowledge communities is essential to understanding progress, and might be centrally important to scientific researchers and firms investing in R&D (Jones, 2021). We compare the knowledge content

and the rhetorical content of each of these knowledge communities to see if we can find a basis for differential success. We find that the way knowledge communities learn from particular knowledge as a foundation and the way they generate knowledge as rhetoric to introduce new ideas have opposite tendencies, and are under strong and constant tension. In successful knowledge communities, knowledge content is drawn from a broad range of sources and is extremely flexible. In contrast, successful communities tend to use very similar rhetoric, their rhetorical content tends to remain central over time, and it tends not to be unique or esoteric when compared to the vocabulary of other communities.

In Chapter 3, we conduct a type of meta-analysis on corporate strategy itself, the very field we are hoping to challenge (Upham et al., 2010a). If we can show that even researchers obsessed with organizations are themselves often driven (at least in part) by their intense membership to knowledge communities, we hope to have accomplished a sort of research Judo. We focus on how researchers position their ideas in and between knowledge communities to maximize their impact and success. Management strategists tend to share their knowledge with each other and with practitioners quite freely. In this field, ideas are hard to patent, and the social uptake of ideas and their success are directly intertwined (Abrahamson, 1996; Eccles, Nohria, & Berkley, 1992; Meyer, 1999; Ponzi, 2002).

Using a clustering methodology developed for this work, we look at the structure of the field's intellectual content and how publishing authors interact in patterns with one another (Hotelling, 1929). We study how the impact of published papers differs depending on how the authors position their papers within and between knowledge communities, which are defined intellectual communities.

We find that it is significantly beneficial for new knowledge to be a part of a knowledge community and also that, within a knowledge community, new knowledge has much more impact if it is in the intellectual semi-periphery, rather than at its periphery or center. This implies a curvilinear relationship within a knowledge community between a work's range of citations and impact. Further, we find some conflicting evidence about whether working in multiple schools of thought results in higher impact. We also show that the conclusions we draw in this chapter and the previous one, each in a different context, apply to both of the datasets. We hope this helps generalize our conclusions and, statistically at least, increase the robustness of our findings.

In Chapter 4, where we use emerging clusters in the hard sciences and basic research, we begin to explore the mechanisms by which knowledge changes over time, with the foundations of innovation and technology: basic research (Small & Upham, 2009; Upham & Small, 2010). We study the most technical fields where so many new scientific ideas percolate first. While this research has the potential to be commercialized (leading to the value of proprietary knowledge), at least in the U.S.A. and Europe as dominant geographies for patents and scientific research, research is still largely federally funded and, overall, the framework of shared basic science still dominates (Gulovsen, 2019; Ma & Uzzi, 2018; Murray & Stern, 2005).

We determine the areas of science that have the most cohesive high-intensity innovation, which we call "emerging fronts." For these fronts, we interview experts in the scientific community to verify our quantitative findings. We begin to consider how a citation structure within scientific research can reveal the processes of change within intellectual communities and how new research emerges into prominence from existing knowledge, potentially to break away and become a distinct field of knowledge itself.

Emerging fronts of research represent the very beginnings of new and cohesive micro-revolutions in science, something only a few have attempted to study systematically at this stage. We select highly cited and cohesive sets of research, which represent the emergence of fresh and potentially important innovations. The insights gained by our qualitative interviews are used to test the success of our method and to contribute to future attempts to identify emerging fronts of research.

The research sources used by these fronts can help us identify patterns. We look at cross-disciplinary research to see if there is a pattern of increased interdisciplinary research among these high-impact clusters over the past five years. Academic research in the sciences has been shown to not only generate important “basic” research that is used by scientists and researchers, but also to have significant spillover into the private sector.

A NEW DAWN

Technology has long been a driver of change and disruption, and the rise of decentralized, dynamic knowledge communities is no exception. The Internet allows sharing and collaboration on a scale never before seen; it should be of no surprise therefore that it has impacted the process of innovation itself. While geographic technology clusters, like Silicon Valley and research universities, like MIT or Harvard, play an enormous part in furthering knowledge, it is also important to understand how they are drawing on and interacting with broad research efforts in other fields, regions, and institutions. Blockchain technology means that decentralized communities and DAOs may not require intermediation by formal organizations. Innovation has gone global; a new invisible set of informal networks have formed that will drive its pace and direction. Clustering technologies provide a window into this phenomenon that could prove to be valuable in understanding how innovation happens in a modern, connected world—and how to potentially help it flourish.

Most of the scarce existing literature on knowledge communities (often described in different terms, such as paradigms, clusters, or invisible colleges) has focused on describing the phenomenon rather than how it works and how it emerges and evolves, or why it is accelerating. Little research has looked at what features in a knowledge community increase innovation rates over time. In this book, we begin this task by first constructing a robust clustering methodology and then trying to tease out and test the characteristics that allow knowledge communities to succeed. We then study what incentive structures exist for individuals to join and grow their respective knowledge communities. Lastly, we explore how powerful new ideas are “born” in emerging fronts and what characteristics allow them to gain intellectual coherence in their early days, eventually potentially becoming larger knowledge communities.

Innovation and technology are working on fundamentally different principles than before innovators, scientists, and others could share information so broadly and instantaneously with others they did not know and might never meet. This has resulted in the need to rethink the importance of formal firm boundaries, in favor of a network perspective of knowledge.

As is so often the case, the whole “tree” may have been contained within the tiny “acorn” and inevitable from the start—the very first internet in the world, the ARPANET, was formed to link Pentagon-funded research institutions and to accelerate research collaboration. We are excited about how a better understanding of decentralized knowledge communities can accelerate innovation and technology.