The Spatial and Institutional Factors
of Knowledge Production at MIT

By
Matthew Claudel

B.A. Architecture
Yale University, 2013

SUBMITTED TO THE DEPARTMENT OF URBAN STUDIES & PLANNING

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE
IN URBAN STUDIES & PLANNING

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2016

Submitted by: **Signature redacted**
Matthew Claudel, Department of Urban Studies & Planning, Master of Science

Certified by: **Signature redacted**
Carlo Ratti, Professor of the Practice; Director Senseable City Lab,
MIT Department of Urban Studies and Planning, Thesis Supervisor

Accepted by: **Signature redacted**
P. Christopher Zegras, Associate Professor, Master in City Planning Committee Chair,
Department of Urban Studies and Planning

1

The Spatial and Institutional Factors
of Knowledge Production at MIT

by
Matthew Claudel

Abstract

Academic research is increasingly cross-disciplinary and collaborative – around the globe and within institutions. In this context, what is the role and relevance of the campus? I examine the scholarly output and collaboration patterns of the Massachusetts Institute of Technology as they pertain to its interrelated organizational structures, including institutional affiliation and spatial configuration, over a 10-year time span. There are four significant results: 1. diverging trends in the composition of collaborative teams over time (size, faculty versus non-faculty, etc.) between papers and patents; 2. patterns of cross-building and cross-disciplinary collaboration are substantively different between papers and patents; 3. a network topology and community structure that reveals spatial versus institutional collaboration trends; and 4. a persistent relationship between proximity and collaboration, well fit with an exponential decay model, that is consistent for papers, patents, and for specifically cross-disciplinary work. These insights contribute an architectural dimension to the field of scientometrics, taking a first step toward empirical foundations for applied institutional policy and spatial planning.

Thesis Supervisor: Carlo Ratti
Professor of the Practice; Director Senseable City Lab, MIT Department of Urban Studies and Planning

Thesis Supervisor: Fiona Murray
Associate Dean for Innovation, Co-Director MIT Innovation Initiative, William Porter Professor of Entrepreneurship, Faculty Director Martin Trust Center for MIT Entrepreneurship, Faculty Director Legatum Center, Assistant Professor, Sloan School of Management

Acknowledgements

Outline

Abstract
Acknowledgements
Table of Contents

1. Introduction

1a. Campus space and Collaboration at the Research Institute

The practices of academic research are deeply tied to physical space and to institutional organization (Knorr-Cetina 1981; Owen-Smith and Powell 2004), yet the sited collaboration networks within communities are not well understood (Wineman 2008, 2014). Examining the interactions of individual-level, building-level and department-level attributes with respect to scholarly output provides insight into the complex processes of knowledge creation. This is particularly salient in the contemporary research environment – one that is increasingly cross-disciplinary, collaborative, and reliant on digital communication tools (Jones 2008; Bikard 2015).

I focus this case study on the Massachusetts Institute of Technology, a community of scholars organized in various interrelated departments, labs, centers and institutes.[1] Some of these bring researchers together in a dedicated building,[2] some share a building between research groups,[3] while others are institutional frameworks that link people who are separated by great spatial or intellectual distance.[4] Considering several large datasets (faculty attributes, campus facilities, publications and patents), I implement tools of geospatial analysis, network science, and statistics in an empirical study of scholarly activity over the ten-year period from 2004 to 2014, drawing on an emerging set of socio-spatial and network-based methods (Boschma 2005; Catalini 2012; Kabo 2014; Wineman 2014).

In addition to a rich portrait of scholarship at the institute, there are four significant results from this analysis: 1. diverging trends in the composition of collaborative teams over time (size, faculty versus non-faculty, etc.) between papers and patents; 2. patterns of cross-building and cross-disciplinary collaboration are substantively different between papers and patents; 3. a network topology and community structure that reveals spatial versus institutional collaboration trends; and 4. a persistent relationship between proximity and collaboration, well fit with an exponential decay model, that is consistent for papers, patents, and for specifically cross-disciplinary work. Finally, I identify the possibility of a 'distance limit' for collaboration, and propose strategies to better evaluate complex evolving networks.

This hybrid analytical approach to the interactions between space, collaboration, and academic performance is of theoretical value as a methodology, contributing a spatial dimension to the established field of scientometrics (for example, Newman 2001a, b; Girvan 2002; Barabási 2002; Borner 2004; Bikard 2015) and a network-science approach to space syntax and planning (for example, Hillier 1984; Peponis 1998, 2007). Furthermore, a multivariate analysis of scientific production and collaboration provides a better quantitative understanding of the qualitatively understood activity of the institute. Both the pedagogical and architectural foundations of MIT emphasize cross-disciplinarity and a merger of theory and practice. MIT founder William Barton Rogers forged a novel academic model in his polytechnic for applied science, and the architect of MIT's present Cambridge facility, William Bosworth, explicitly engaged collaboration with his innovative design for an interconnected campus (Appendix A). Today, in the context of new buildings, research groups, initiatives, and accelerators with a stated ambition to support collaboration, it is crucial to build a body of empirical knowledge and spatial practices that reinforce the core innovation-driven ethos of MIT.

---

[1] http://web.mit.edu/industry/departments-labs.html
[2] Koch Institute
[3] Department of Brain and Cognitive Sciences
[4] MIT Energy Initiative

I propose that a deeper knowledge of the specific academic and relational phenomena on this campus could, in the future, be distilled into a set of frameworks that may have a practical impact on the MIT community, offering a data-driven foundation for applied institutional policy, architectural design and campus planning. These frameworks may then be generalizable to a number of analogous contexts – those that can also be framed as innovation ecosystems. On both sides of the academic/non-academic divide, I suggest that a new approach to socio-spatial organization may ultimately enrich the design and operation of innovation ecosystems, promoting knowledge creation, collaboration and commercialization of research. This conceptually positions a dynamic planning concept – envisioned by MIT urban scientist Kevin Lynch – as part of an answer to the great mandate for innovation issued by Joseph Alois Schumpeter in 1932. In the future, we may operationalize a rich understanding of collaboration processes to design active spaces that reduce the indeterminacy of collaborative innovation.

> "*Novelty* constitutes what we here understand as *development*... The triad *'indeterminacy, novelty, leap'* remains unconquerable. [It is] the major problem of today's theoretical economics."
> J.A. Schumpeter, *Development*, 1932.

1b. Literature Review

Science-of-Science and the Triumph of the Team
In recent years, standardized and accessible databases of scholarly material (for example, papers and patents) as well as the increase in computational power to analyze data, have enabled a new field of scientometrics: the application of analytical tools to understand the patterns, processes and dynamics science itself.

Among the strongest conclusions drawn by research in this field is the trend toward collaboration: contemporary science is dominated by teams (Porter 2009). A study of 19.9 million papers and 2.1 million patents created over the course of fifty years shows that co-authorship in scientific publication has increased across nearly all disciplines, and that collaboration produces work with a higher impact (Wuchty 2007). A similar trend is evident in patenting. At the time of the Patent Act in 1952, the lone inventor dominated science: 82% of patents were single authored, and only 3% had three or more inventors. Almost sixty years later, in 2011, 32% of patents were issued to single authors, while 43% had three or more on the collaborative team – and the trend is rising (Crouch 2011). Between the years 1975 and 2000, the average number of inventors per patent in the US rose from 1.7 to 2.3 – consistent across all observed subfields of science – and team patents were invariably more highly cited (Wuchty 2007). By 2010, the average team size had risen to 2.84, with the 90th percentile represented by teams of up to five co-inventors (Rosenberg 2010; Huang 2012). In addition to larger team sizes, science across many fields is becoming more interdisciplinary, drawing on a greater variety of skills and expertise (Porter 2009). It is now common for teams to span many different institutions and to cross national boundaries (Wuchty 2008; Huang 2012; Hsiehchen 2015).

If knowledge is created by groups, the research team is the "fundamental unit in the social organization of science," today more than ever. Revealing the dynamics of team formation and evolution is a crucial aspect of scientometrics (Latour 1987; Milojević 2014). "Understanding the distribution of the number of coauthors in a publication is of fundamental importance, as it is one of the most basic distributions that underpin our notions of scientific collaboration and the concept of 'team science.'" (Milojević 2014). The importance of recognizing and interpreting these characteristics has led to the emergence of a 'science of *team* science' (Wuchty 2007), a sub-field that addresses questions of group size, formation and success.

The rise of collaboration and team-based research has several factors, including individual specialization, intensified telecommunication, and a general trend toward complexity across disciplines. The global scientific community is increasingly a *network* rather than a constellation of isolated inventors. It follows that if an individual can be represented as a node who is connected to others – through co-authorship, co-invention, or even citation, proximity and institutional affiliation – mapping scholarly collaboration through network science reveals the web of connections that constitute globalized academia.

A rich subfield of scientometrics studies these scholarly networks, through the application of models, graph theory and mathematical analysis to bibliometric data (for example, Newman 2001a, b 2004; Barabási 2002, Girvan 2002). Network analysis has revealed defining characteristics of scientific production, most notably: the academic community exhibits 'small world' properties; clustering coefficients vary widely across fields, but most have an 'invisible college' of interconnected scientists; a high assortivity coefficient shows that the most gregarious scientists tend to be connected with each other (Newman 2004), among many other insights. Network analysis can reveal collaboration dynamics across geographic scales and at different levels of disciplinary aggregation, from the building to the campus to the international research community.

Communication Through Co-location in the Research Lab
As a function of internal and external social network effects, successful teams evolve toward a size that is large enough to enable specialization and effective division of labor among teammates, but small enough to avoid overwhelming costs of group coordination (Guimerà 2005). Specifically, diversity can spark creativity while simultaneously presenting the risk of conflict and misunderstanding (Larson 1996). Team science, then, is contingent on communication. Prior to digital telecommunications, the mechanism of collaborative information transfer was simply face-to-face communication, within the laboratory or near to its physical boundaries (Allen 1986). To support collaborative communication as a function of proximity (and for a number of associated factors, including the use of technical equipment), research and its practitioners have historically been organized into teams, working in labs, sited on campuses. These conditions allow for the explicit or implicit establishment of scientific knowledge and practices, and the rapid dissemination of ideas (Latour 1979, 1987; Knorr-Cetina 1981).

The lab, as a social and spatial unit, is a crucial enabler of research, through facilitating conversation and the development of shared intellectual premises. Intuitively, co-location in a lab means spatial proximity, but it also entails a number of concomitant 'proximities' – for example, the development of shared research procedures, language, or values. That is, there are many facets of proximity, and many distinct but interrelated phenomena and effects. The concept of 'proximity' has been treated in several disciplines, but it is a term without a common definition. To clarify what is identified as "ambiguity of the proximity concept," across scientific fields (and even within disciplinary literatures) Knoben (2006) presents a comprehensive multivariate classification that includes cognitive, organizational, social, institutional and geographical proximity – a productive synthesis of theoretical and empirical approaches to understanding proximity effects. The present analysis focuses generally on geospatial proximity, which is shown in many cases and contexts to be an important factor of innovation processes (Oerlemans 2005)

A Finer Grain of Spatial Analysis
In addition to social composition, a consideration of the architecture itself provides a richer understanding of the complex dynamics of collaborative research. Allen (1977) postulated an inverse relationship between the likelihood of communication between employees and the distance between their offices, with a limit of 50 meters as a benchmark for regular (weekly) technical communication. Subsequent work introduced more sophisticated metrics for spatial proximity – multi-variable configurations and building-scale systems (labs, floors, hallways, organization, paths, and architectural character), rather than simply the linear distance between two points. In fact, quantifying the built environment has become a field in its own right, broadly known as 'space syntax' (Hillier 1984). For the present analysis, I apply a simple metric of spatial proximity, but acknowledge the value of implementing a more sophisticated characterization of the physical environment.

The limited case studies that have leveraged a richer set of empirical tools for space syntax broadly confirm that physical configuration has a direct impact on the frequency of contact between co-workers in a corporate environment, and in turn, on sustained work-related communication deemed 'useful' (Penn 1999). Factors of co-location that support collaboration include awareness, body language, auxiliary media, and unplanned informal conversation (Fussell 2001). Evaluation of these so-called 'situated social networks' with respect to spatial design and productivity reveals a correlation between spatial proximity, social connection, hub spaces, and collaborative output in the context of corporate research (Penn 1999). Pilot studies in a university have examined similar social and spatial networks among scientists in specific buildings, finding that the formation of new collaborations and the resulting success in grant applications are significantly and positively influenced by the co-location of researchers (Kabo 2014, 2015; Wineman 2014). Being retrospective, however, these studies may be biased by endogeneity – researchers are not randomly positioned, and as such, collaboration may not be purely a function of proximity.

Any collaborative research process is an extended engagement between its participants, with distinct phases and protocols. Intuitively, proximity has different effects on each of these specific segments. Certain tasks – generating ideas, structuring work flow – are benefitted by co-location, while others are not – for example, completing independent technical work (Fussell 2001; Shah 2012). Strategically planned phases of temporary proximity, then, may be a useful approach for successful collaborative work. There is a line of research that addresses these targeted re-combinations, finding that co-location can be a productive tool in the ideation and execution phases of a research process, contingent upon a high degree of organizational and cognitive proximity (Torre 2004; Boschma 2005; Knoben 2006; Torre 2009; Balland 2015). In my discussion of future work, I suggest an analysis of discrete shifts in proximity and spatial configuration as a way of controlling several confounding variables.

The practice of sequencing phases of in-person meeting and remote communication is changing with advances in digital technology (Finholt 1997). Analyses of communication patterns have tested the effects of space and medium in digital collaboration, specifically, the use of face-to-face communication versus telecommunication technology in academic research environments (Kraut 1988; Baym 2004), in large corporations (Penn 1999), and in controlled test situations (González-Ibáñez 2012). Generally, the frequency of email communication correlates directly with telephone and face-to-face communication, in both social and professional relationships.

Despite that digital telecommunication may be more effective than face-to-face communication in many cases (Townsend 1998; Fussell 2001; Gilson 2015), spatial proximity remains relevant for three reasons: serendipity, initiation, and information quality. Co-location is an important factor in initiating contact (Penn 1999), turning a good idea into a collaboration, and diffusing knowledge (Allen 1984; Stuart 2010). Telecommunication cannot, in common academic practice, provide the spark of serendipity – the unexpected exchange of information that leads to novel ideas. Spatial configuration provides opportunities for unplanned encounters between people who are affiliated with disparate parts of an organization (Wineman 2014) and contributes to the formation of social networks that are crucial to innovation (Wineman 2009). Nearby colleagues tend to collaborate more frequently (Fussell 2002), and, in both cases of organizational similarity and dissimilarity (i.e. two researchers being in the same department or in different departments), physical proximity can induce collaboration between people who may not otherwise work together. Indeed, it has been argued that, as it defines serendipitous communication, space plays a key role in the broad advancement of science (Hillier 1991).

Scales of Collaborative Activity, from the Desk to the Globe
A. Microgeographies: The Desk
The configuration of personal workspaces – at the scale of a desk in a room – certainly define social interactions. Those configurations may also define the *types* of ideas that are collaboratively generated, through increasing exploration during low opportunity-cost time, and enabling accelerated iteration during early or uncertain phases of collaboration (Catalini 2012; Miron-Spektor 2011). Yet in the case of workplace microgeography, it is often difficult to identify cause and effect: whether researchers work nearby because they collaborate, or if they collaborate because they work nearby. To address concerns of endogeneity in microgeographic organization, studies have been designed to exploit a random exogenous factor – for example, the unexpected spatial reconfiguration of researchers. Over the course of a move to new premises, Peponis (2007) conducts survey-based social and spatial network analysis, finding that layout can contribute to frequency and volume of communication. A similar analytical design by Catalini (focusing on reconfigurations during building closure for asbestos screening), demonstrates that, beyond simply provoking communication, shifts in proximity lead to increased experimentation and eventually to stronger 'breakthrough ideas' (Catalini 2012).

B. Hubs: The Building
The majority of the literature reviewed previously – that addressing lab configurations, office proximity, space syntax, collaboration and re-configuration – falls within the architectural scale.

C. Ecosystems: The District
Although it has no discrete boundary or categorical definition, the 'district' is perhaps the most relevant scale for innovation processes with respect to commercialization. There is a rich literature from urban planning, management, organization science and economics, as well as an emerging field of innovation science, that considers this so-called 'ecosystem' or 'cluster' – an area smaller than a city, and more heterogeneous than a campus. Topical literatures have emerged around specific terms and definitions, but all address this district scale. These include 'Marshallian industrial districts,' 'innovative milieus,' 'new industrial spaces,' 'innovation networks,' 'regional network evolution,' and most recently, the 'ecosystem-approach' (Porter 1998, 2001). Each has a unique classification – for example the 'cluster' definition is a geographic concentration of interconnected companies and associated institutions in a particular field, linked by commonalities and complementarities (Porter 1998; Glaeser 2010; Delgado 2012).

This work – and the management science literature in general – primarily addresses market-related characteristics and effects of proximity. Porter's seminal work on entrepreneurial clusters, for example, describes inter-firm relationships, and evaluates such phenomena as the competitive transfer of financial and human capital. In contrast, the present analysis focuses on interrelations, knowledge-creation and technology transfer outside of (or prior to) commercialization, but nonetheless accesses the literature describing general effects of geographical and institutional proximities (campus space and departments, respectively).

The *innovation ecosystem*, as defined in the MIT Stakeholder Model is a *vibrant, co-located agglomeration of high-growth potential firms and related stakeholders, an interconnected set of people, resources and the physical environment that provides the context for innovation-driven enterprises to start, grow and scale*. The model describes place-based interrelations of five key stakeholders: entrepreneurs; universities or research institutes; government; corporations; and risk capital.

D. Agglomerations: The City
Understanding collaboration and innovation at the urban scale (and the inter-urban, national and global scales) demands fundamentally different tools and procedures. It is at the metropolitan level that *data* becomes *big data*, beyond this threshold it can only be approached through a specialized set of statistical, social, mathematical and computational analysis techniques – from the collection, management and processing of data, to the application of statistical models, to means for addressing concerns of subject privacy, and even to visual representation and dissemination.

Intuitively, metropolitan-scale conclusions can be drawn using this approach. Stern and Guzman, for example, interact a number of publicly available data points related to new firms (including name, keywords, intellectual property, legal status, etc.) and construct an ex-ante model of entrepreneurial quality: the statistical likelihood of success outcomes, defined as acquisition or IPO. The resulting Regional Entrepreneurship Cohort Potential Index (RECPI) can be trained on metropolitan regions to estimate the impact of a city itself on its firms: the Regional Entrepreneurship Acceleration Index (REAI) measures the performance over time as the ratio between the statistical success potential of a cohort of firms and their observed success rate in that city. This tool has been used to evaluate and compare cities across the United States, finding characteristic patterns of entrepreneurship (Stern 2015).

Even more generally, certain urban effects related to collaboration and knowledge output are quantifiable through a big data analysis approach. For example, a number of metropolitan phenomena – including patent filings, for example – scale superlinearly with city size (Bettencourt 2007). An empirical model proposed by researchers at the Santa Fe Institute is consistent across cities, time periods and nations. Derivative growth equations evidence a number of urban taxonomies, for example, the distinction between urban growth fueled by entrepreneurial innovation versus growth through economies of scale, as measured by wealth and patents (Bettencourt 2010). Trends in collaborative invention are significantly spatialized as well. Large metropolitan areas tend to have a higher number of patents and larger team sizes: an average of greater than three co-inventors. The most highly productive metropolitan areas have over four inventors per patent (Rothwell 2013). In the case of inter-urban and global collaborations, a characteristic distance – unique to specific fields – can be quantified (Tijssen 2011).

These, and a number of related studies, access scientometric observables (for example, collaboration, publication density, co-inventor distance, and others) and apply macro-level tools of urban science, delivering valuable insights at a low level of spatial granularity. For the purposes of the present research, city-scale dynamics are important for contextualization, but specific results are not easily transferrable to the campus or building scale.

Defining the Research Institution

There are deep entanglements between all elements of an innovation ecosystem, however, the present study is narrowly concerned with the research institute and its innovation capacity (which may ultimately serve as a foundation for broader ecosystemic research). An understanding of the dynamics of the institute itself is nonetheless valuable outside of academia, as a function of innovation and entrepreneurship: research universities are a significant driver of global economic growth, through educating a highly skilled segment of the workforce. MIT alone has an extraordinary economic impact – companies founded by graduates generate an estimated $1.9 trillion in global revenue, roughly equivalent to the GDP of the 10th largest economy in the world.[5] Even more salient to the present analysis, universities and research institutions produce cutting edge knowledge (specifically, as patents), which exits into corporations or new innovation-driven ventures that are poised for scale (Murray 2015).

A spectrum of organizations, including public and private universities, technical institutes, sponsored research centers, corporate research and independent labs have varying configurations with respect to knowledge production. Among institutions of higher education, the most important distinction is between the *university* and the *research institute*. The designation of "university" has different meanings around the world, and no officially standardized definition even within the United States. University status is determined at the state-level – Massachusetts, for example, grants the title only to institutions that offer two or more doctoral degrees.[6] The title is not definitive or categorical, but I propose a classification based on process and output: the *research institute* engages in knowledge production, both as pure science and as practical applications, through licensing intellectual property and engaging in industry partnerships. Drawing from the *Frascati Manual*,[7] I further categorize the practice of research into three different activities: basic research, applied research and experimental development.

[1] Murray, Roberts, Kim. "Entrepreneurship and Innovation at MIT: Continuing Global Growth and Impact." 2015
[6] Massachusetts Board of Education. "Degree-granting Regulations for Independent Institutions of Higher Education," 2011
[7] OECD, "Frascati Manual: The Measurement of Scientific, Technological and Innovation Activities," 2015

1. Basic research is either experimental or theoretical, with the intent of advancing knowledge and/or establishing an empirical foundation for further research, without an explicit non-academic application.
2. Applied research is similar in its process and scope, but structured and directed toward future developments with a specific non-academic application or practical objective.
3. Experimental development draws on both basic and applied research, as well as real-world analysis and contextualization. It is directed to producing new materials, systems, products, processes or devices, and commonly involves testing or demonstration.

The present study evaluates MIT – a technical institute that engages in all three: basic research, applied research and experimental development. This is conducted both as pure science and as industry-sponsored research, and is supplemented by the institute's curriculum in the social sciences and the arts (see Appendix A). Intellectual material produced at MIT exits the institute through a number of different pathways, including scholarly publication, popular media, licensed intellectual property, built projects, and human capital.

The Porous Walls of the Institution: Managing Technology Transfer
A categorization of research is particularly relevant with respect to incentives and procedures for publication, implementation and commercialization. Technology transfer is an important economic driver, and foregrounds the role of intellectual property rights policies (Rosenberg 1994). The pathway from research to commercialization was fundamentally redefined in 1980 with the Bayh-Dole Act (or Patent and Trademark Amendments Act) that is in place today. The act systematized and unified patent policy across federal agencies, and allowed for universities and research institutes to hold patents that result from government funded research. This legal measure shifted incentives for researchers of all levels to patent and license through the institution, and and for the organization itself to engage in joint research ventures or enable spinoff startup companies. An increase in the commercialization rate of university-based technology following 1980 can be attributed in some part to the Bayh-Dole Act.

A more explicit approach to technology licensing has given rise to the 'economics of intellectual property' at universities and research institutions (Link 2003). Many institutions now have technology transfer offices that serve to manage the exit dynamics of scientific research. Although terms are unique to each organization, in most cases there are clear motivations for faculty to disclose and file patents through the institution, at various stages of technical resolution, from proof of concept to product (Jensen 2003).

At MIT, the Technology Licensing Office (TLO) assists the passage of research from the lab to the market, through patenting and licensing. The TLO interfaces with large corporations, SMEs and startups – brokering the legal use of patented MIT research – and generally assists students, faculty and staff with their pursuit of innovation-driven work. The TLO files patents with USPTO, and has worked with an average of 20-27 new firms per year from 2000 to 2014.

Systematic technology transfer procedures enable a more accurate understanding of knowledge spillover and the regional impact of research institutions. Prior to the Bayh-Dole Act, proximity-driven pipelines from research to commercialization were diverse and ad-hoc, and often outside the institution's purview. It is now possible, however, to use the linking patent to empirically distinguish formal university-firm partnerships from 'pure' knowledge spillovers (Monjon 2003). More generally, the cluster and ecosystem literatures stand to benefit greatly from the increasing precision of observable knowledge objects through the tools of network and urban sciences.

Observing Scholarship: Faculty, Papers and Patents

Proximity-enabled communication – whether in the lab, the building, the campus, the institution, or the cluster – is a crucial enabler of scientific knowledge creation. *Place* is at the core of the university model throughout history. Yet for intellectual material to be recognized, validated, and built upon or commercialized, it must be codified and disseminated to the global scientific community, specifically, as papers and patents: observable knowledge objects. Although these documents are not comprehensively representative of intellectual activity at a given institution (and their role varies dramatically across disciplines), they nonetheless offer a systematic proxy for scholarly output. Bibliometric data enables empirical analysis of impact, affiliation, collaboration, and citation over time.

For the purposes of this study, I consider papers published in peer-reviewed journals, and patents filed through the MIT Technology Licensing Office, but a portrait of scholarly activity at MIT should be contextualized by broader trends in scientific production. In the case of paper publishing by the global scientific community, individual productivity (on both single-author and collaborative observations) has a fat-tailed distribution: a small number of scientists produce a very large number of papers, and this has been consistent over time (from Lotka's Law of Scientific Productivity, proposed in 1926, to Newman's network analysis in 2001). Synthesizing several estimations of bibliometric data, Borner proposed a general model for the growth of scientific publishing as well as co-authorship and citation trends, successfully validated against twenty years of publication data from the Proceedings of the National Academy of Sciences (Borner 2004). The global rate of increase in scholarly publication (estimated as the number of cited papers that are subsequently cited) is approximately 8% (Bornmann 2014).

The World Intellectual Property Organization (WIPO) estimates a similar growth for intellectual property. Excepting a downturn during the global economic crisis in the years 2008 and 2009, the rate of increase in patenting is now approximately 9.2%.[8] College, university, and institute patents represent between 4.2 – 4.7% of U.S. nongovernmental patents, of which Massachusetts Institute of Technology holds an annual average of 4.2% (of all U.S. patents granted to U.S. academic institutions).[9]

Time Delays in Academic Output

There is a significant time delay in the course academic production from idea to publication. The sequences for papers and patents are categorically different, resulting in a significant discrepancy between overall time delay (a discrepancy generative of certain experimental structures, e.g. Murray and Stern, 2007). Although there is variation across fields, papers are accepted, on average, 6.4 months after submission, and are published 5.8 months after acceptance – the total average time delay is 12.2 months from submission to publication in a journal (Björk 2013). The patenting process is substantively different, and tends to vary widely between different contexts. In the case of MIT, an individual approaches the TLO with a patent application, and works with the office to revise the document. A formal application is filed by TLO at the US Patent and Trademark Office between 0 and 5 months later, depending on revisions. There is then a period of 24-36 months before review by a patent examiner, followed by a granting decision period of 12-24 months. The total time elapse from submission to granting is an average of 2.5-5 years.[10]

---

[8] WIPO. World Intellectual Property Report 2015 - Breakthrough Innovation and Economic Growth. 2015

[9] National Science Foundation Report, from http://www.nsf.gov/statistics/seind12/c5/c5s4.htm

[10] Direct communication with MIT Technology Licensing Office, 04.20.2016

Institutional Affiliation

There is no standardized taxonomy of academic affiliation across universities and research institutions – the titles "professor" or "lecturer," for example, may entail different roles in different organizations. MIT recognizes a number of affiliations that designate very different research, publication, patenting and compensation protocols. The table below shows titles (MIT Affiliations) and those roles categorized strictly as faculty.[11]

MIT Faculty	MIT Affiliations
Assistant Professor (Tenure-track)	Professor of the Practice, Adjunct Professor (or Associate)
Associate Professor (Tenure-track)	Visiting Professor (or Associate or Assistant)
Associate Professor (Tenure)	Instructor (or Technical)
Professor	Lecturer (or Senior or Visiting or Honorary)
	Affiliated Faculty
	Fellow (or Post-Doctoral)
	Research Scientist (or Researcher)
	Student

---

[11] MIT Institutional Policies, from: http://web.mit.edu/policies/index.html

1c. Hypotheses

I posit that the spatial organization of MIT faculty is an important factor of the knowledge creation process, and examine the ways in which this is manifest in scholarly output, including rate, volume, authorship and collaboration patterns of publishing and patenting. To approach this question, I investigate a number of specific sub-hypotheses:

1. There are inherent differences in output (rate, volume and authorship structure) by department, and between papers and patents. Succinctly, that scholarly artifacts evidence disparities across fields of science and modes of scientific production.
2. Building heterogeneity is positively correlated with cross-disciplinary collaboration (internal and external to the building).
3. Faculty tend to collaborate within buildings, and with faculty who have a higher degree of spatial proximity (that there is a negative correlation between distance and probability of collaboration) even if collaborators are in different departments.
4. The community structure of the co-authorship network has a higher degree of similarity to departments or initiatives, and the community structure of the co-invention network has a higher degree of similarity to buildings.
5. Instability of spatial organization (a re-location event) is correlated with an increase in scholarly output.

Generally, I suggest that certain buildings function as collaborative hubs – particularly active, densely and heterogeneously populated spaces. These play an important role in initiating contact among researchers, specifically those from different disciplines who would not otherwise meet. As such, researchers sited in hub spaces will exhibit a higher instance of cross-disciplinary collaboration.

Similarly, because spatial proximity – regardless of specific building location – is an important determinant of regular communication and a crucial element of successful collaborative work, I propose that it will be associated with higher incidence of collaboration, regardless of the specific building. These characteristic socio-spatial patterns will be different between papers and patents, as a result of the inherent disparity between collaboration processes (related to the use of technical equipment, research lab structure, and the fact of publishable novelty), disciplinary fields, and academic norms for attribution between papers and patents.

2. Data and Method

Method

There are three original databases underlying the present study: 1. Directory; 2. Publication; 3. Patent. These three databases are linked using the MIT Identification Number: a unique 9-digit numerical value assigned to each MIT affiliate, which persists through changes in affiliation over time. The linked data is limited, by time and scope, to ensure precision, as discussed in detail below. The truncated data is then processed to derive second-order variables (e.g. a faculty member's ratio of papers to patents), as well as aggregated (e.g. to the department or building level) and linked to auxiliary data (e.g. mapped to building footprints to calculate geodetic distance between faculty). What follows is a summary of the original data (A1-3), the processing method (B1-2), and the final dataset (C). The final dataset spans 2004 – 2014 and encompasses 40,358 papers, 2,350 patents, and 1,057 MIT faculty. Datasets and key attributes are summarized in Table 1.

A1. Original Data: Directory

Directory information is from the MIT Directory (accessed in the MIT Data Warehouse, via the Office of Institutional Research). The full dataset includes all MIT affiliates, including students, researchers, fellows, tenure track and non-tenure track faculty, etc. The MIT directory contains location-attributes for offices (building, floor and room), as well as affiliation attributes (school, department, or lab), but the information is not recorded for every person (see B2). The MIT Data Warehouse commits a regular 'freeze' of the data, such that time-dependent information is captured (e.g. changes in affiliation or title).

A2. Original Data: Publication

Publication information is from a comprehensive list aggregated by Academic Analytics, a non-institutionally affiliated data analytics company. Academic Analytics aggregates publication data from scholarly journals, for the purposes of evaluation, strategic decision-making, and benchmarking in universities. Publications in the Academic Analytics database have been consistently filtered with a proprietary algorithm to match individual authors, and the company's total database includes over 270,000 individuals from over 10,000 departments in 385 universities.

MIT-specific data has been acquired by the Office of Institutional Research (OIR) and is stored in the MIT Data Warehouse. The dataset includes papers published by MIT-affiliated individuals in peer-reviewed journals with DOI number identifiers, as well as date, and authorship. Data is stored such that a paper is associated with an MIT ID, and the same paper is repeated if there are more than one MIT authors on the paper. There are 65,536 MIT authorship instances (with repeated DOIs) and 52,511 papers (with unique DOIs) spanning the years 1959 to 2015. In this dataset, there are 1,440 total MIT authors (faculty and non-faculty).

A3. Original Data: Patent

Patent information is from the database of patent applications filed through the MIT Technology Licensing Office (TLO) and aggregated in the MIT Data Warehouse (via the OIR). The dataset includes a unique identifier (Patent Serial Number) for each patent, as well as date of filing, MIT ID of primary inventor, and total authorship (all inventors listed on the patent). There are 7,344 patents, and 5,487 MIT inventors (faculty and non-faculty) over the years 1938 to 2015.

Pre-Processing Data

Through pre-processing the total databases, several necessary limitations and filters have been introduced to increase accuracy of the data under consideration. These filters are: B1. Time Frame (2004 – 2014); B2. Affiliation (Tenure and Tenure-track Faculty) to arrive at a final dataset, C.

B1. Pre-Processing: Time Frame

The temporal limitation has two reasons. First, there was a sharp increase in the number of recorded papers in the year 2003 (Figure A1), which may have resulted from Academic Analytics' migration to a digital database for content management. Although there are a small number of papers recorded as early as 1959, trends and attributes are not consistent before and after the shift. Secondly, the practice of regular data 'freezes' committed by OIR / MIT Data Warehouse began in 2004. In the interest of understanding changes over time (e.g. hiring or departing, affiliation, promotion, office location) rather than working under the assumption that a faculty member's current status is true of past years, I consider only this range, limiting the data to the time frame 2004-2014.

B2. Pre-Processing: Affiliation

There are numerous modes of affiliation to MIT, many of them involved with the processes of research and the production of academic output. For example, an undergraduate student may participate in groundbreaking research and be listed as an author on a paper with her Principal Investigator. A comprehensive portrait of scholarly activity at MIT would capture these myriad interactions.

However, several attributes are consistent only for a subset of the MIT community – the result both of structural factors and data management. The former is intuitive: in the case of the UROP, she would not have a designated office location or even a departmental affiliation. The same is true, to varying levels of consistency, up to the level of post-doctoral fellows, lecturers, or professors of the practice. Furthermore, the data 'freeze' only captures attributes for certain levels of affiliation. The subset with full consistency is tenure and tenure track faculty, who are necessarily affiliated to a department and located in an office. As such, the most accurate and consistent directory includes only faculty who are tenure or tenure track.

C. Data Summary

The dataset under consideration for the remainder of the study is characterized in Table 1.

	Papers	Patents
Observations	40,358	2,350
Avg Annual Number of Observations	3,668.91	213.64
Collaborative Observations with >1 MIT faculty member	6,414	454
Number of Faculty	1,035	359

Table 1. Summary of data after pre-processing. This dataset, including MIT tenure and tenure-track faculty between the years 2004 and 2014, will be used for the remainder of the study. The table lists unique documents as *Observations*; the average annual output of the MIT community over the time frame; observations with more than one MIT faculty involved; total number of publishing and patenting faculty (not mutually exclusive).

Spatial Data

Spatial data describing the MIT campus is joined from several different sources. Building shapefiles (in GIS format) are from the Department of MIT Facilities and overlaid on publicly available map tiles (from OpenStreetMaps via CartoDB). Building-level data (area, facility use codes, office numbers and floors) are from the MIT Data Warehouse. The databases are linked using building code (e.g. Buiding E70; Building 9).

3. Descriptive Statistics

3a. Community-level

Output per Person

At the individual level, faculty who publish created an average of 29.2 papers (σ 24.9) during the ten-year time span, and faculty who patent created an average of 5.77 patents (σ 4.30). A detailed summary of individual output for the MIT faculty community is presented in Table 2 – information that serves to contextualize my interpretation output trends at higher levels of aggregation and further analysis. For example, faculty who are outliers in productivity may occupy an important place in the total network – they may serve as nodes with high betweenness (Newman 2001a, b; Barabási 2002) or represent structural holes in community clusters that import and export good ideas (Burt 2004). Network topology and communities are further evaluated in Section 6.

Annual Output

Between the years 2004 and 2014, overall paper and patent output from MIT has increased. Publication rate has remained relatively constant over time, while patents demonstrate more variability. Assuming a linear growth trend (R^2 = 0.77), there is an average annual increase of 23.84 patents per year from the entire MIT community. Paper output exhibits a stable growth trend (R^2 = 0.96), with an average annual increase of 131.58 papers per year (Figure 1). These trends are statistically significant over the 99% confidence interval.
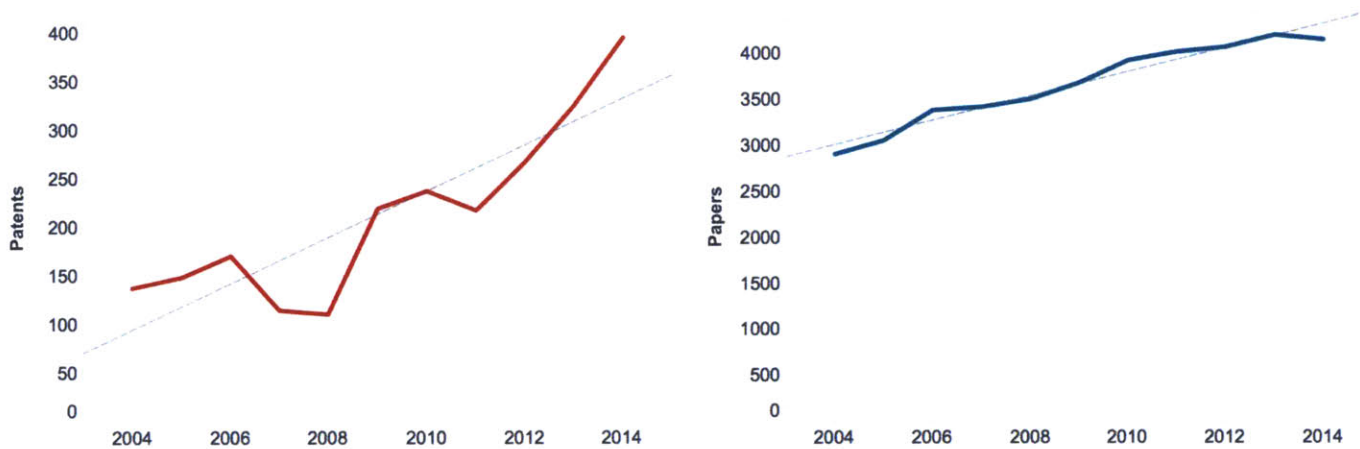

Figure 1. Annual patent and paper output by MIT faculty, between 2004 and 2014, with trend lines describing growth. Patents are fit with R^2 = 0.77 p-value of 0.0004 and t-value of 5.45 and papers are fit with R^2 = 0.96 p-value <0.0001 and t-value 14.12

Comparing Paper and Patent Output

The MIT community has produced an increasing number of papers and patents per year, but absolute output numbers are very different between the two output types. To compare the trends observed between papers and patents, I calculate the relative frequency of observations per year for both output types – considering total output of each as 1 and annual output for each year as a fraction of the total (Figure A3). Patents appear to be increasing at a rate that is better fit with an exponential curve, however I calculate the linear fit to facilitate comparison. On average, the patent output of the total MIT faculty community has been increasing more quickly than paper output, at a relative rate of about 1% per year versus .3% per year, respectively (in total, output has been increasing annually by an average of 131.58 for papers and 23.84 for patents)

3b. Department-level

Departments
In considering the entire MIT academic community (rather than a domain-specific subset) it is necessary to acknowledge variations between disciplines, particularly with respect to the publication of papers versus patents. Some fields produce knowledge that is inherently un-patentable, but those fields are no less active. A prolific faculty member of Mechanical Engineering, for example, will have a very different resume than her colleagues in Political Science. A comparative analysis of scholarly output should account for – or at a minimum, distinguish between – fields and sub-fields of science (a common practice in scientometrics, for example: Newman 2001, 2004; Breschi 2007; Björk 2013; and others).

The following results are not a relative comparison between departments, but a summary of characteristic trends by discipline, which serves to inform the interpretation of subsequent results. As expected, there are significantly different patterns of paper and patent output by department. I characterize paper and patent output by 33 departments and research groups in Appendix Table 1, and summarize the highest producing departments for each publication type in Table 2. In the following section, I evaluate the same characteristics per year – for example, papers per person per year, at the department level.

Department	Patent to Paper Ratio	Number of Faculty Authors	Total Papers / Person	Total Papers	Average Annual Increase	Intra-MIT Collab.	Number of Faculty Inventors	Total Patents / Person	Total Patents	Average Annual Increase	Intra-MIT Collab.
Biology	0.036	60	59.28	3557	14.52	20.4%	33	3.79	125	0.34	21.6%
Chemical Engineering	0.132	35	80.17	2806	14.47	25.3%	26	14.96	389	5.03	34.2%
Chemistry	0.096	34	79.12	2690	2.87	15.3%	22	11.68	257	1.7	30.0%
Electrical Engineering & Computer Science	0.069	135	58.47	7893	13.95	25.2%	90	6.2	558	6.25	30.1%
Materials Science and Engineering	0.117	35	73.00	2555	7.76	25.1%	28	10.57	296	1.03	30.7%
Mechanical Engineering	0.098	79	48.06	3797	26.51	17.8%	54	7.26	392	5.09	18.9%
Physics	0.019	82	54.52	4471	3.85	29.8%	6	14.5	87	0.25	78.2%
Program in Media Arts & Sciences	0.135	22	47.41	1043	7.72	12.7%	17	8.88	151	2.2	5.3%

Table 2. Patent and paper output by department, for the top six departments in each category. *Patent to Paper Ratio* represents the total number of patents from a department divided by the total number of papers, such that 0 represents a department without patents and 1 a department without papers. *Intra-MIT Collaboration* is calculated as an average of the number of documents with another MIT faculty member vs. the total number of documents, such that values closer to zero represent a lower rate of intra-MIT collaboration. *Number of Faculty* enumerates the total number of publishing faculty from a given department, during the time frame. *Average Annual Increase* represents the slope of the trend line in output per year, a measure of the average increase in output year-to-year, by department.
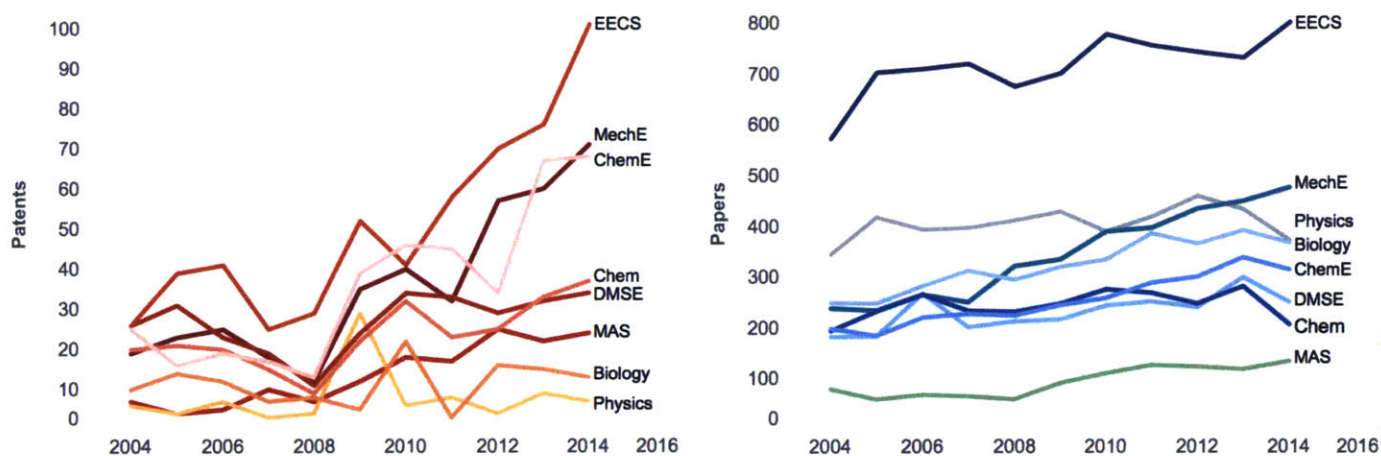
Figure 2. Patent and paper output per department per year. Paper output remains fairly constant at the department level, and as a trend across all departments. Conversely, some departments exhibit a strong increase in patenting while others do not – there is a notable increase in the disparity of output over the course of the decade (excepting a nearly universal dip in 2008)

The top producing departments follow very different patterns over time, for both papers and patents. The former maintains a fairly constant upward trend, within individual departments and across the entire publishing MIT community, while patenting is more variable. Some departments exhibit a strong increase while others do not, resulting in a divergence over time (excepting a nearly universal dip in 2008). The top five patenting departments have very similar output in 2004 (a range of less than ten patents per year), and vary dramatically over the following decade (to a range of over 80 patents per year).

To better characterize the decrease in 2008 and ensure that it is not an inconsistency in the data, I plot the same curve for comparable institutions – Stanford, CalTech and the UC system, the top patenting institutions in the United States (Figure A2). A similar pattern appears, particularly in the UC system. Further research into the procedures and incentives for patenting suggests that this aberration may have been related to a bill passed by the United States Senate that changed the funding structure of the USPTO, and caused a transition to a 'second pair of eyes' review system, resulting in a sharp decrease in patent grants across the country.

At the department level, there are meaningful trends in academic fields – for example, in 2012, over 50 MIT physicists (including 6 faculty members) were involved in the breakthrough discovery of the Higgs Boson particle,[1] in collaboration with the European Organization for Nuclear Research (CERN). The discovery is regarded as one of the greatest scientific advances of the year (or decade), and resulted in a number of associated publications. It is also possible to observe, in a course way, discrepancy between paper and patent sequences – specifically, EECS exhibits a peak in paper output around 2010, and only in the years following (2011 – 2014) a steep growth in patent output.

---

[1] MIT News Office. "CERN announces observation of new particle consistent with long-sought Higgs boson. Dozens of MIT researchers contribute to breakthrough in particle physics." July 4, 2012

3c. Individual-level

The summary of department-level trends is not intended as a comparison between those departments. The very fact of publishing a paper or filing a patent is substantively different from one discipline to another. Furthermore, departments vary significantly in size – from Electrical Engineering & Computer Science, with 135 faculty members, to the Music & Theatre Arts Section, with only 1. Table 3 summarizes output per person per year, aggregated to the department level. Annual averages account only for years during which a specific faculty member was affiliated with a specific department.

Unevenness of Productivity
Department-level output is not evenly distributed among faculty members. In some departments, a small number of faculty produce a disproportionate share of the academic work, while in others, faculty tend to publish comparable numbers of papers and patents. A measure of the unevenness of productivity informs my interpretation of department-level results.

Departments can be characterized by the variance in output at the faculty level. For each faculty member, I evaluate the total number of observations during the time frame, accounting for each individual's number of publishing years. In this way, it is possible to control for the inherent bias toward faculty who have been publishing at MIT for a longer period of time. Variance is given by:

$$s^2 = \frac{\Sigma(X - M)^2}{N - 1}$$

M = mean of department output (per person, per year)
N = number of faculty in the department
X = faculty output (average per year)

I also test a characterization of faculty output per department using Shannon Entropy of Information function (see Section 5b) as a comparison, but find that variance is a better representation of disparity in faculty output. For example, a department with a cluster of faculty who have very high output and a cluster with very low output (evenly distributed across the mean) will be calculated with high entropy, despite a large variance. Departments with the greatest and least variance in both output types are summarized in Table 3 (full list in Appendix Table 2).

Department	Avg. Patent per Person per Year	Paper Variance	Rank Order	Avg. Paper per Person per Year	Patent Variance	Rank Order
Chemistry	5.22	51.2	4	4.44	290	1
Chemical Engineering	4.44	75.7	3	2.61	27.0	2
Physics	7.59	148.8	2	0.49	4.8	8
Biological Engineering	5.76	16.4	13	1.40	25.1	3
Mechanical Engineering	6.68	27.0	9	1.39	14.9	4
Materials Science & Engineering	5.13	50.1	5	2.28	10.4	5
Mathematics	1.94	6.7	17	0.10	0.20	19
Architecture	1.37	1.1	23	0.21	0.20	18
Earth, Atmospheric & Planetary Sciences	3.34	12.2	14	0.07	0.10	21
Ocean Engineering	0.03	0	32	0.66	1.10	14

Table 3. The unevenness of output per faculty member in each department. In some departments, a relatively small number of people generate the majority of the output, while in others, output is evenly distributed among faculty. *Average Annual Output Per Person* accounts for the number of years that a faculty member was affiliated with MIT (to control for the inherent bias toward those who have spent a longer time at the institute).

Papers and Patents at the Department-level and Individual-Level

As shown in Table 2, many departments produce both papers and patents (as do individual faculty members!) but, intuitively, departments tend to have a bias toward one output type – for example, Mechanical Engineering and Chemical Engineering emphasize patenting, while the Physics and Biology Departments focus on publishing, relative to the total MIT community. Notably, Electrical Engineering & Computer Science has a high paper *and* patent output.

But this trend could be attributed to department size – more publications would almost necessarily be produced by a larger number of faculty members, and EECS is one of the largest. I consider paper and patent output per person to control for department size and identify individual-level characteristics of different scholarly disciplines. These trends are remarkably different from department-level activity: faculty in the Physics department characteristically focus more on patents than papers, as compared with the average ratio. Faculty in the Chemical Engineering department are the outlier in paper and patent output per person.

I also acknowledge that there are some disciplines whose activity is captured neither by papers nor patents – for example, Architecture, in which faculty produce books, design projects, or built work. The discussion and conclusion suggests future work that may better capture this activity.
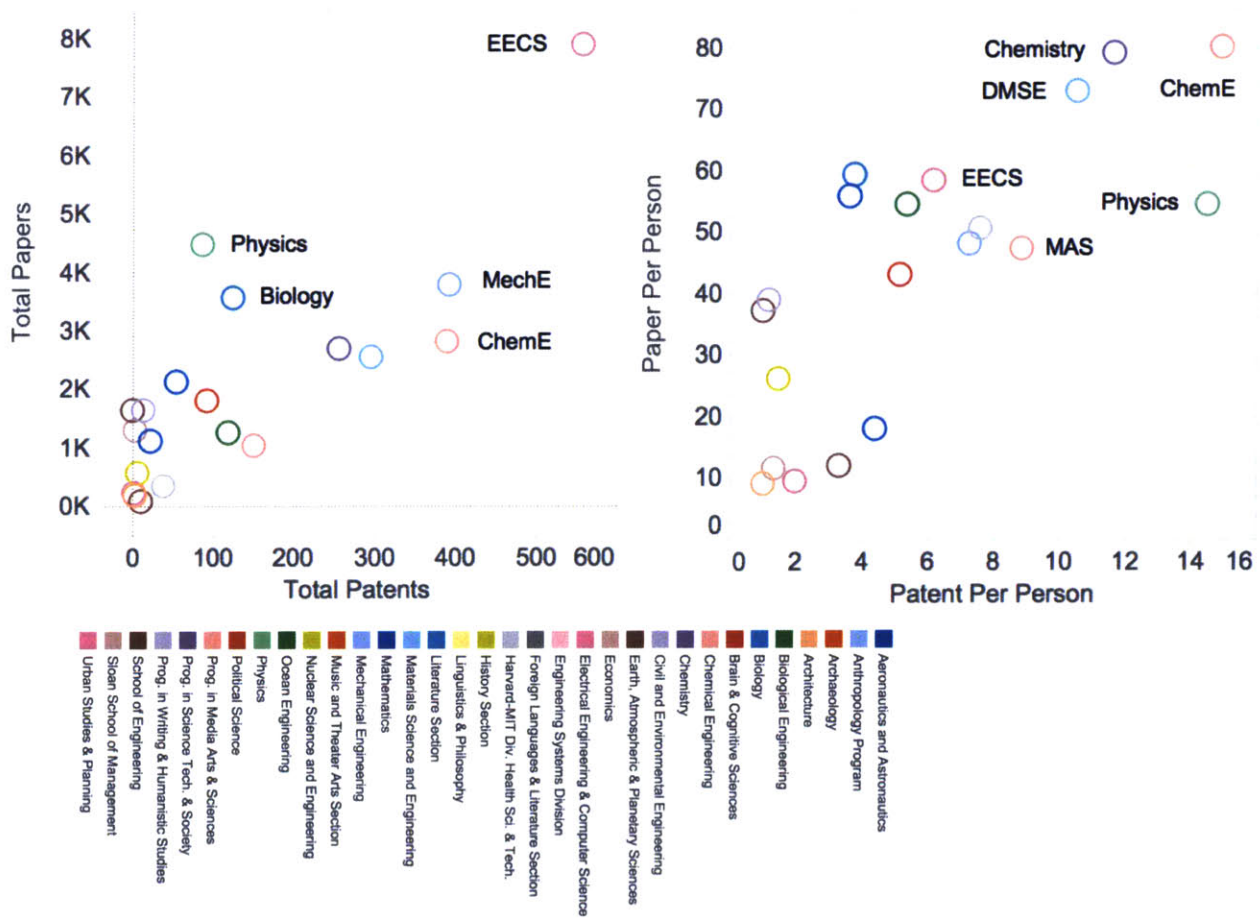

Figure 3. The ratio of paper vs patent output, aggregated to the department and individual levels. Each circle represents a department, and output per department / per person determines position along the axes. Many departments tend to have a bias toward one output type, but a notable group produce high relative numbers of both papers and patents.

3d. Building-level

Which spaces on campus are hotspots of scholarly productivity? Each building at MIT is uniquely configured – some correspond one-to-one with a department, others are shared among several departments, and most departments are spread across several buildings. I begin by examining the distribution of scholarly output across the campus, aggregated to the building level.

Mapping Building-Level Data
Figure 4 shows a choropleth map of campus with buildings coded according to aggregate volume of scholarly output for the decade. Table 4 shows the top five buildings for each publication type, as well as the output per person and the collaboration ratio – a ratio of the number of documents with another MIT faculty member versus the number without. Maps of output per person are shown in Figure A4.

Building	Papers per Building	Intra-MIT Collaboration	Papers per Person	Patents per Building	Intra-MIT Collaboration	Patents per Person
32	4166	27.8%	110.0	140	30.7%	3.68
3	2570	19.3%	93.0	290	21.4%	8.29
13	2370	27.5%	78.9	278	45.7%	12.64
36	1949	19.6%	78.4	171	19.3%	7.77
6	1704	20.5%	77.4	157	42.0%	13.08
66	1662	28.8%	71.8	169	46.2%	8.45
46	1607	13.6%	67.2	69	7.2%	3.63
54	1372	12.2%	60.5	1	100%	1
68	1330	24.1%	59.4	22	9.1%	2.44
33	1285	16.8%	59.2	40	25.0%	4.44

Table 4. Patent and paper output per building, for the top ten buildings, during the entire time frame. *Output per Person* represents the average individual output aggregated to the building level. *Intra-MIT Collaboration* is calculated as the number of documents with another MIT faculty member versus the total number of documents (aggregated to the building level), such that values closer to zero represent a lower rate of intra-MIT collaboration.
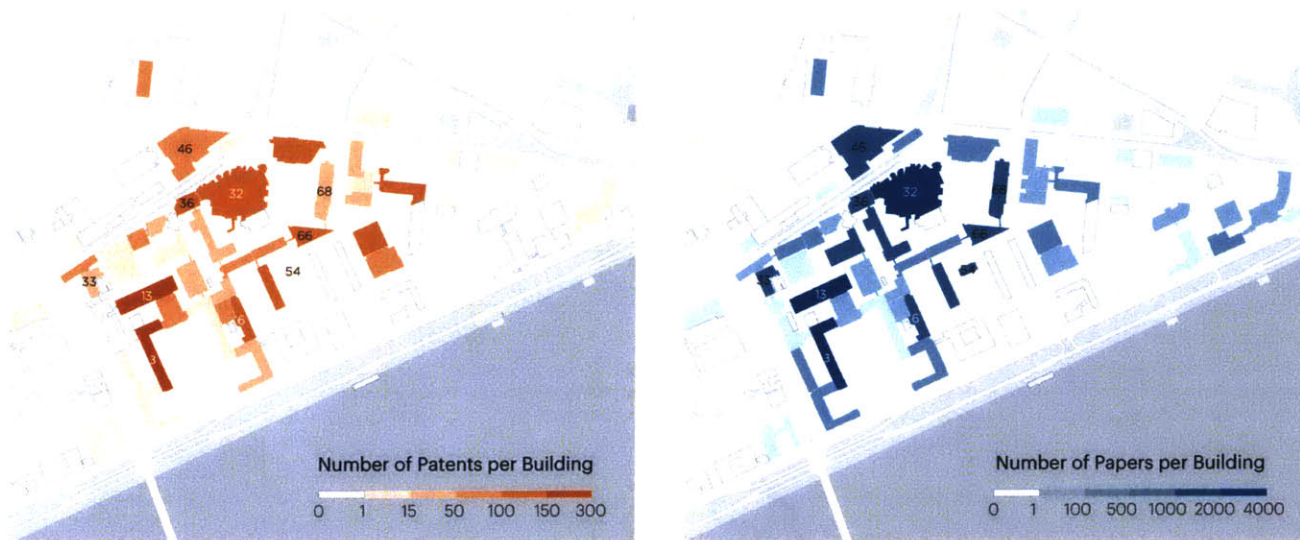

Figure 4. Patent and paper output per building between 2004 and 2014. On these choropleth maps, buildings are color coded by output volume and labeled with their name (the facility code). Colors are assigned using a Jencks algorithm with five buckets.

Beyond the overall characteristics, how have these buildings performed year by year? Considering annual volume can evidence specific events (changes in funding, closures or developments in the field, for example) that may have impacted the rate and volume of knowledge output. Figure 5 shows building-level trends for paper and patent output from 2004 to 2014, counting only those observations that were created in a given building during a given year.
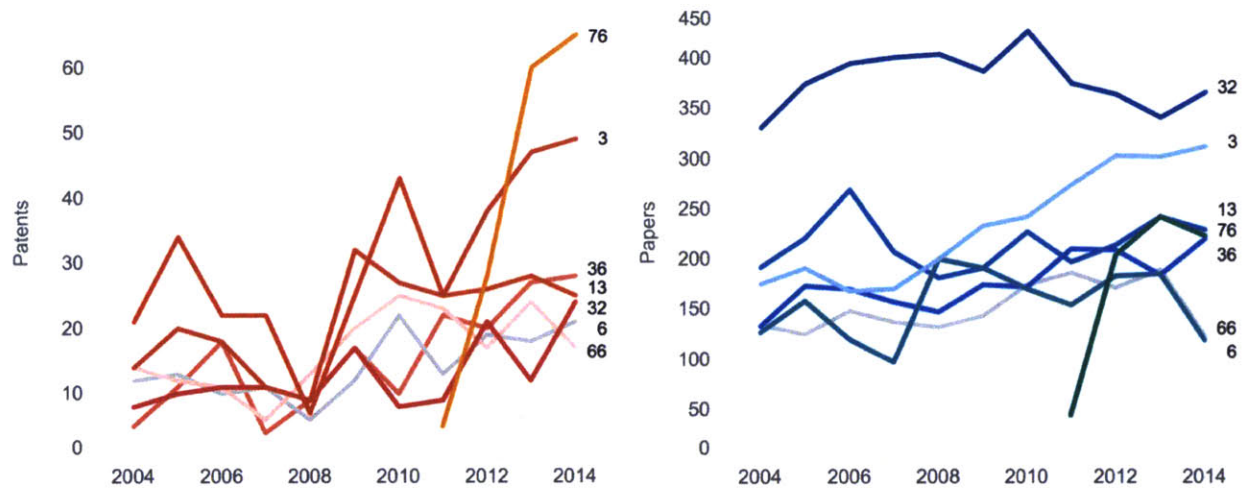

Figure 5. Patent and paper output per building per year, from 2004 to 2014. This represents output from faculty during the year they were sited in a particular building – accounting for spatial relocations. Building 76, the Koch Institute, was opened in 2010 and by 2013 had become the top patenting building.

Building 13 exhibits a rise in paper output in 2006, and a corresponding peak in patents in 2010. Generally, this evidences the time delay between output types, but a consideration of those specific documents reveals that the peak is largely tied to the discovery of graphene – a breakthrough that was made in 2004 in the United Kingdom, and expanded soon after by researchers at MIT. Several faculty members in Building 13 were key players in this research.[2] Building 3 demonstrated rapid growth in patent output after 2011. Is it unique or different? Did it experience change?

Building 76 certainly did – better known as the Koch Institute for Integrative Cancer Research, the center was initiated through a $100 million grant from David H. Koch in October 2007, with a mission to bring together distinct approaches to cancer research. This first shift connected researchers who were formerly disparate in terms of affiliation, and provided funding for research partnerships. In December 2010, the new building was inaugurated, offering specialized equipment and facilities, and serving as a nexus for cross-disciplinary work. The intra-building collaboration effects are further analyzed in Section 5a.

More broadly, many MIT buildings host a wide variety of departments, labs and research groups – Building 3, for example, hosts over sixteen labs and research groups (Section 5b). My hypothesis is that such spaces will facilitate and promote collaborative activity – given that proximity and co-location are vital to research, particularly across disciplinary boundaries – prompting a further analysis of collaboration and building diversity.

---

[2] Halber, Deborah. "Nanoscience rising up to meet energy challenge, Dresselhaus says," MIT News Office. April 11, 2006

4. Collaboration

The production of knowledge is becoming increasingly collaborative. The broader scientific community has shown a trend toward collaborative work – particularly internationally – over several decades: the rise of 'team science' is replacing the traditional image of the lone genius. The average number of inventors per patent is growing – across all subfields of science – and team patents are invariably more highly cited. The same trends hold true for co-authorship and citation (Wuchty 2007). To better understand the rise of collaboration and validate its consistency in the MIT academic community, I evaluate co-authorship and co-invention at the total, departmental, building, and individual level. The analysis draws on a growing field of network-based scientometrics coupled with spatial analysis.

Overall Co-authorship and Co-invention
The current analysis distinguishes between individuals with and without MIT faculty status. Non-MIT faculty may include collaborators at MIT who are not designated as faculty members (for example, researchers, visiting lecturers, students, postdoctoral fellows) or collaborators who are not affiliated with MIT (for example, a faculty member of another university). There are an average of 4.00 total co-inventors on patents filed through MIT – greater than the national average of approximately 2.84 – indicating a propensity to work together in larger teams. The average number of MIT faculty inventors per patent is 1.45 (σ 0.36) and has increased over the time period by an average of 0.011 faculty inventors per year. There are an average of 3.95 total authors on papers associated with MIT, of which an average 1.23 are MIT faculty co-authors (σ 0.25), a number that has also experienced a slight increase during the time period, by approximately 0.028 MIT faculty authors per year (Figure A5). While patents and papers exhibit a similar average number of total collaborators – 4.00 and 3.95 respectively – the ratios of faculty to non-faculty collaborators are different.

To better illustrate team composition, I calculate a simple ratio of MIT faculty to non-MIT faculty collaborators for all MIT papers and patents per year. A value of 1 represents a document with *only* MIT faculty collaborators, and a value approaching 0 would indicate *only* non-MIT faculty collaborators (0 is impossible, as our dataset is based on MIT faculty output). Figure A6 shows that the collaboration ratios of papers and patents – rates that have remained fairly constant over time.

Department	Faculty Inventors *Mean 1.45 (σ 0.35)*	Total Inventors *Mean 4.11 (σ 0.84)*	Faculty : Non Faculty Ratio	Faculty Authors *Mean 1.23 (σ 0.24)*	Total Authors *Mean 3.41 (σ 0.77)*	Faculty : Non Faculty Ratio
Physics	2.19*	4.92	0.45	2.32*	4.81*	0.48
Nuclear Science & Engineering	1.66	3.33	0.50	1.20	4.20*	0.29
Biological Engineering	1.24	4.58	0.27	1.43	4.26*	0.34
Chemical Engineering	1.47	4.59	0.32	1.38	4.09	0.34
Electrical Engineering & Computer Science	1.49	3.74	0.40	1.43	3.83	0.37
Mathematics	2.27*	6.05*	0.38	1.29	3.29	0.39
Earth, Atmosphere & Planetary Sciences	2.00*	6.00*	0.33	1.19	3.89	0.31
Architecture	1.33	5.00*	0.27	1.26	3.35	0.38

* Indicates values greater or lesser than one standard deviation from the mean, outside the interquartile range.

Table 5. Co-invention and co-authorship at the department level. *Faculty* designates MIT faculty members, as opposed to *Total*, which represents the number of cited authors on a document. *Faculty : Non-Faculty Ratio* is the relationship between these two values, for each output type.

Department-level Collaboration

Sub-fields of science have very different characteristics of co-authorship and co-invention, in addition to the observed differences in volume and rate of scholarly output by department. To better understand these differences at MIT, I first evaluate team sizes by department – the number of authors or inventors per document aggregated to the department level. The average number of MIT faculty co-inventors aggregated to the department level is 3.6 and the average number of total co-inventors is 4.1 (standard deviation of 0.9 and 0.8 respectively). Mathematics, EAPS and Architecture are outliers in patenting team size.

As compared to patents, intra-faculty publishing follows a consistent pattern in author numbers – the standard deviation is relatively small. There is, however, a notable exception in the Physics department, where there are an average of 3.64 faculty co-authors. These characteristics prompt a deeper analysis of building-level co-invention and co-authorship dynamics. Is the tendency to collaborate with other MIT faculty related to spatial configuration? Is it related to output type? Are there differences related to cross-disciplinary collaboration (between different departments) and intra-building collaboration?

5. Spatial Analysis

5a. Intra-building and intra-department collaboration

Faculty members in the same department are often sited in the same building. In order to better understand the effect of co-location as distinct from co-affiliation, I examine the relationship between departmental organization and shared physical space. There are 3,886 unique co-author pairs and 538 unique co-inventor pairs (repeating individual faculty who collaborate with a number of different faculty members). Each collaborating faculty pair is assigned a binary value for co-location and co-affiliation (1 for same, 0 for different). Furthermore, I find that for both papers and patents, collaborators who are in the same building have a higher than average link weight – a measurement of collaboration intensity (Section 6a):

	Co-Author Pairs	Avg. Link Weight	Co-Inventor Pairs	Avg. Link Weight
Total Network	3,886	0.29	538	0.21
Intra-Building	1,232 (31.7%)	0.32	142 (26.4%)	0.22
Intra-Department	1,892 (48.6%)	0.31	238 (44.2%)	0.21
Intra-Bldg and Dept	1,124 (28.9%)	0.32	134 (24.9%)	0.23

A similar metric can be calculated for individual papers and patents with more than one MIT faculty collaborator. Each observation is assigned a value for building and department diversity among its contributors, using the Simpson Diversity Index. The index is a quantitative measure that reflects the number of different types (such as species) there are in a complete dataset, and simultaneously takes into account how evenly the basic entities (such as individuals) are distributed among groups of those types. The numerical value increases both when the number of types increases and when evenness increases. For a given number of types, the value of a co-affiliation index (C) is maximized when all types are equally abundant. In this case, the number of species $i = 1, 2, ..., d$ are defined by the total number of unique departments or buildings, and the number of different unique collaborators n_i who belong to a given type i. If an observation is characterized by N authors and d types, the diversity index is given by:

$$C = \frac{1}{N(N-1)} \sum_{i=1}^{d} n_i(n_i - 1)$$

Where the co-affiliation value (C) of a single observation is a function of total number of contributors (N), from a total number of departments (d), and (n_i) number of contributors from each of those departments (i). Values are on a scale from 0 to 1, representing least to greatest co-affiliation or co-location among collaborators.

Patterns of intra-building and intra-department collaboration between papers and patents are remarkably different (Figure 6). For papers, the rate of collaboration within a buildings and departments is fairly consistent and relatively high, and the disparity between the two could be interpreted as the incongruity between buildings and departments (some departments are in several buildings, and some buildings host several departments). Conversely, patents are less internally collaborative overall, but rates are highly variable and increasing. Figure 6 shows the annual average co-affiliation index per observation, for those observations that are co-authored with another MIT faculty member (of a total 710 authors and 214 inventors collaborating internally within MIT).
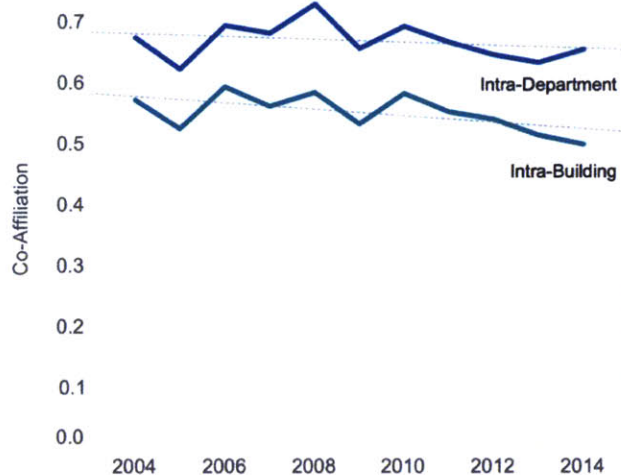
	Patent	Paper
Intra-Building		
Average	0.23	0.55
σ	0.07	0.02
Intra-Department		
Average	0.42	0.67
σ	0.09	0.02

Figure 6. Each observation that has more than one MIT faculty collaborator is assigned a value (from 0 to 1) such that higher values indicate a greater incidence of co-affiliation, for either a building or department. Notably, rates of collaboration within buildings and departments have increased for patents, and decreased for papers. The latter is far more variable, and exhibited an intra-departmental peak in 2009, followed by an intra-building peak in 2012.

The rates of intra-building and intra-departmental collaboration on papers are roughly parallel over time. Patents, however, are more varied – intra-departmental collaboration spiked in 2009, and intra-building collaboration in 2012. As discussed in Section 3b, the USPTO's shift to a 'second pair of eyes' system caused a universal dip in patenting activity around 2008. It is possible that, for those patents that were nonetheless granted in 2008-2009, there was an increased incentive to co-invent with faculty members from the same department.

In the case of intra-building co-invention, Figure 5 may shed light on the reasons for changes in activity. The plot of overall output suggests that the Koch Institute for Integrative Cancer Research (Building 76) may have been an important factor in the 2012 spike. Funding was announced in 2007, and the center's new building opened in 2010. Given that patents at MIT take an average of 2.5-5 years to be granted (Section 1b), one could read the pattern observed in Figure 6 as a result, in part, of the contribution of the Koch Institute. Since its first patent in 2011, Building 76 quickly rose to be the top inventing building (Figure 5). Furthermore, the building has the highest rate of overall intra-MIT co-authorship, defined as the percentage of total publications that are with another MIT faculty member (Table 6).

5b. Building-level

Diverse Spaces

Not every building is occupied in the same way – some are allocated entirely to a single department or lab, while others host a diverse group of faculty from various disciplines. Buildings that have a higher degree of disciplinary heterogeneity may be areas of intellectual cross pollination. Heterogeneity of faculty affiliations can be quantified at the building level, defined as a function of the number of faculty and the number of different departments represented by those faculty members, using Shannon's entropy of information equation.[3]

The measure was originally proposed to quantify entropy in strings of text: the greater the difference in letters, and the more equal their proportional abundances in the string of interest, the more difficult it is to correctly predict which letter will be the next one in the string. The Shannon equation is applicable more broadly to quantify balance of information content, accounting for variety and quantity. In the case of MIT campus, I apply the following function to characterize buildings. Given a building k its heterogeneity index λ_k is then defined as:

$$\lambda_k = -\sum_{i=1}^{D_k} p_i \ln p_i$$

Where D_k = number of different departments in building k, and p_i is the fraction of people who belong to the department i. It is defined as:

$$p_i = \frac{N_{ik}}{N_k}$$

Where N_{ik} is the number of people that belong to department i in the building k, while N_k are the total number of people working in building k. There is a distribution of values between 0 (Building E62, which hosts 93 faculty members, all in 1 department, and 2.54 (Building 4, which hosts 8 faculty members in 4 departments)(Table 6). Using a correlation coefficient analysis, I examine whether there is a statistically significant relationship between building heterogeneity and academic output, but, contrary to my hypothesis, find none. Across the total MIT community, diverse buildings do not necessarily lead to greater individual paper or patent output, as illustrated in Figure A9. Figure 7 maps campus buildings, coded by heterogeneity value.

Furthermore, I evaluate the relationship between building heterogeneity and number of collaborators per document. There is no statistically significant relationship for paper author numbers or for inventor numbers. Finally, I test for a correlation between heterogeneity and intra-building collaboration, but find no relationship: none of the trends are within the boundary of statistical significance (Figure A9).

Building 3	
Micro Power and Nanoengineering Research Group	International Consortium for Medical Imaging Tech
Reacting Gas Dynamics Laboratory	Marine Hydrodynamics Lab
Bioinstrumentation Engineering & Microanalysis	Department of Mechanical Engineering
BioInstrumentation Laboratory	Mechatronics Research Laboratory
Cadlab	Nonlinear Systems Laboratory
d'Arbeloff Lab for Information Systems & Technology	Precision Engineering Research Group
Hatsopoulos Microfluids Laboratory	Program in Polymer Science and Technology
History, Theory and Criticism of Architecture and Art	Pappalardo Laboratory

---

[3] Shannon, Claude E. (1948). "A Mathematical Theory of Communication". *Bell System Technical Journal*, 27 (3): 379–423.

Crowded Spaces

Anecdotally, crowded physical spaces for research spur productivity and spark collaborative work – for example, in the history of MIT Building 20.[4] Using MIT facilities data, I sum total lab area and office area and divide by number of faculty who occupy that building – a rough calculation of the amount of active research space per faculty member (in square feet), as a representation of the crowdedness of each building. For each building b we define (c_b) as:

$$c_b = \frac{b_l + b_o}{b_f}$$

Where (b_l) represents the building's total lab area and (b_o) the total office area in square feet, and (b_f) the the total number of faculty located in the building. Figure 8 shows the resulting crowdedness of each building, and values are reported in Table 6. As with heterogeneity, I conduct a correlation coefficient analysis to identify any relationship between crowdedness and paper or patent output, but there is no observable relationship suggesting that less area per faculty member impacts individual output of papers or patents. I examine the relationship between square feet of lab and office space, and number of co-authors per document. Contrary to my hypothesis, there is no discernible trend suggesting that crowded buildings lead to greater numbers of collaborators.
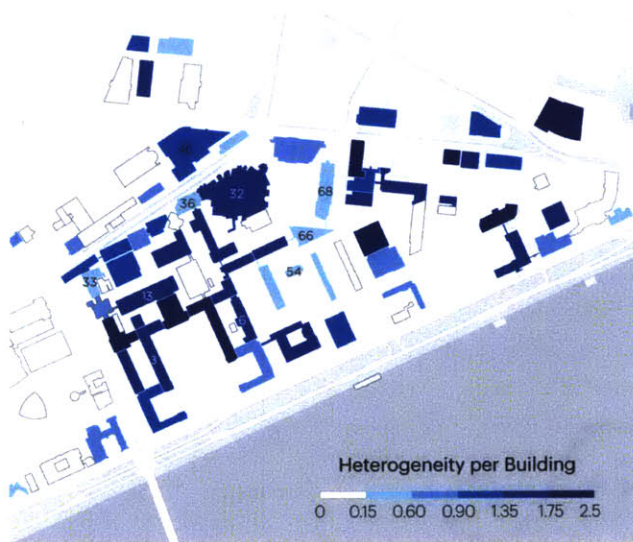

Figure 7. MIT campus buildings, coded according to heterogeneity, as calculated using the Shannon measure of information entropy. This shows variation in faculty departmental affiliations per building. Values range from 0 to 2.5, classified with a Jencks algorithm. Buildings are labeled by name (facility code).
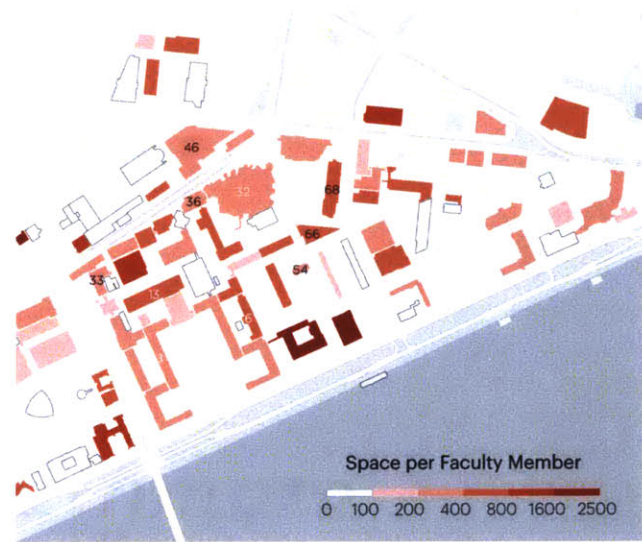
Figure 8. MIT campus buildings, coded according the average total area of lab and office space per faculty member. There is a distribution of values from 145ft^2 to 2,065ft^2 allocated per faculty member. Buildings are labeled by name (facility code).

---

[4] Lehrer, Jonah. "Groupthink: The brainstorming myth," *The New Yorker.* January 30, 2012.

A Diverse Crowd

Some buildings on MIT campus are both crowded and heterogeneous – spaces that could be considered vibrant in a similar way to Building 20. Those with the least space for labs and offices allocated to the most diverse group of faculty are shown in Table 6. However, there is no overall correlation with either paper or patent productivity – all are below the total averages of 6.50 and 0.85, respectively. Furthermore, there is no statistical correlation between diverse, crowded buildings and intra-building collaboration.

Building	Depts.	Shannon Entropy	Average ft.²	Papers			Patents		
				% MIT Collab.	Papers Per Person	Papers	% MIT Collab.	Patents Per Person	Patents
3	9	1.56	340.50	19.3	47.59	2570	21.4	8.29	290
E40	8	2.06	256.80	17.4	13.08	340	-	-	-
32	8	1.71	398.53	27.8	39.68	4166	30.7	3.68	140
16	8	1.44	175.80	26.6	31.92	830	23.9	6.77	88
E25	7	1.64	479.15	16.8	36.33	981	17.5	10.36	114
7	6	2.45	350.54	12.5	4.80	48	-	-	-
1	6	1.28	344.08	12.3	29.85	1015	35.7	1.56	14
76	6	0.89	358.75	31.7	28.52	713	27.2	7.52	158
E38	5	2.28	354.48	21.1	6.33	57	100	1.00	1
10	5	1.99	145.95	19.9	18.17	527	44.6	6.22	56

Table 6. MIT campus building attributes. Although there are no global correlations between heterogeneity and the observed indicators of academic output, there are nonetheless specific correlations between number of departments and output (notably, Buildings 3 and 32). Shown here are the ten buildings with the highest number of departments. *% MIT Collaboration* represents the percentage of observations that are with an MIT faculty collaborator versus the total number from that building.

Despite that there are no global correlations between heterogeneity and output or crowd and output at the campus level, there are nonetheless notable patterns in outliers for disciplinary diversity. The buildings with the largest raw number of departments (for example, Building 3, with nine departments, or Building 32 with eight departments) are also outliers in academic output, total and per person. Simply having a breadth of departments represented in the same building may promote collaboration, particularly across disciplinary boundaries. These buildings could be thought of as hubs on campus. To better assess this hypothesis, I consider *centrality* – a measure of average proximity to everyone else on campus.

5c. Campus-level

Spatial Centrality Index
If space is an important factor of collaboration and proximity facilitates frequent, productive, face to face communication (Kraut 2001; Catalini 2012), then centrally located faculty would be better suited for collaboration. Using the geospatial position of each faculty member's office, I calculate the distance between every pair of faculty. In network terms, this constitutes a complete network, in which each faculty member is a node and the distance between each pair of faculty members is an edge. Averaging the set of all distances for each faculty member is a measure of campus centrality; for example, if a faculty member has the lowest average distance to every other faculty member, she has the highest centrality score. Faculty members can be arranged in rank order, based on the centrality score (lowest to highest average distance to every other faculty member on campus). A faculty member in Building 10, on the infinite corridor in the heart of campus, would have a lower centrality score than a faculty member in Sloan School of Management (E62), at the Eastern edge of campus.

Centrality and Faculty Output, Collaboration, Intra-Building Collaboration
I test the hypothesis that faculty who are more centrally located on campus will have a higher rate of publication or invention. There is a weak relationship between centrality and output, for both papers and patents (Figure A10). Centrality values can be aggregated into bins, such that an average paper and patent output can be calculated for each bin. There is a similar pattern for both papers and patents – a distribution skewed to the right, toward lower centrality, but a large increase at the periphery, at approximately 2500 (Figure A11).

Buildings with a high degree of centrality (defined as the the lowest average distance to every other building) may become hubs for faculty from across campus to meet. Considering building-level centrality, I test the correlation with rate of intra-building collaboration (defined as the ratio of total collaborations to collaborations with another occupant of the building). There is no statistical correlation between building centrality and internal collaboration for patents or publications.

The relationship of central positioning on campus to scholarly output is inconclusive. But centrality, as defined, is a measure of *global* proximity – a faculty member's average spatial relationship to all other faculty members. It is a characteristic of office location, for example, but not an individual-level behavior or attribute. As distinct from 'centrality,' I define 'proximity' as a function of specific faculty collaboration: a characteristic of each pair of co-authors or co-inventors. Proximity is one element of a broader schema for defining the academic community: mapping links (collaboration or spatial relationship) between nodes (faculty or buildings or departments) is, effectively, a network science approach.

6. Networks

In recent decades, the merger of statistics, mathematics, and graph theory has become a *science of networks*. This field offers tools to address a number of diverse situations, from the metabolic function of a single organism to the complex interactions of a biological ecosystem. Networks are a powerful way of understanding urban transportation, collective social organization, software optimization, and more.

Across this spectrum of real-world phenomena, networks exhibit certain properties, including community clustering, small-world structure, betweenness centrality, and preferential attachment. Abstracted theoretical models can shed light on those systems we observe and experience. Newman (2001a, b) demonstrated that scientific collaboration networks, for example, exhibit properties of a so-called "small world network," with short node-to-node distance (average path length) and large clustering coefficient (a strong community structure), a property that was previously defined theoretically by Watts and Strogatz (1998). Intuitively, these two network characteristics result in cluster-based specialization and efficient information transfer across the graph. There are tightly knit groups of academics, and cross-pollinators who span disciplinary boundaries and make connections: the nodes with high betweenness.

The academic community of MIT faculty can be depicted as a network – a group of individual actors who have multiple distinct pairwise relationships to other actors. Using a network science approach, I construct a number of matrices and quantitatively analyze the community. Nodes can represent entities at different levels of aggregation, ranging in size from individual faculty, to labs, buildings, or departments. Links (also known as 'edges') are then generated according to various interactions between actors, for example, co-authorship of papers and patents – as in typical co-authorship networks – or a numerical value representing collaborators' spatial proximity on campus. This is a novel approach, combining collaboration networks and organizational networks with geodetic networks to form a complex multi-layered network that facilitates a comparison of spatial and institutional factors in scientific collaboration.

6a. Multi-layered network definition

Nodes in the multi-layered network represent individual actors – at the finest grain of analysis, these are MIT faculty who publish or patent. Those individuals can be aggregated to varying levels according to categorical attributes, for example, building or department. For every faculty member f, I consider:

Attribute	Description
F	Floor
B	Building
D	Department
b	Publication Index
n	Patent Index

The collaborative graph $G = (N, L)$ is a network composed by N nodes and L links, with different layers of links between network nodes representing specific attributes. An edge layer is generated for each of the following pairwise relationships between nodes i and j

a. Layer 1: *Proximity*

As distinct from 'centrality,' I define 'proximity' as an attribute of faculty collaborator pairs f_{ij}. Using floor (F) and building (B) attributes of each faculty member, a proximity value p can be calculated for every link in the MIT faculty collaboration network. The proximity index is a hierarchical distance function that accounts for both distance between buildings ΔB_{ij} (measured in linear feet between building footprint centroids), and difference in floors ΔF_{ij}

Let i and j be two nodes (faculty) in the MIT collaboration network, and define ΔB_{ij} as the geodesic distance between the centroids of buildings B_i and B_j where i and j are located. Furthermore, let $\Delta F_{ij} = |$ F_i-$F_j|$ be the floor difference between the location of i and j, where F_x is an integer representing the floor in building B_x where x is located. I introduce a scale factor (d) to account for the difference in units of measure between ΔF_{ij} and ΔB_{ij} and calculate d such that:

$$d\,(\max(\Delta F_{\{mn,\dots,ij\}})) < \min(\Delta B_{\{mn,\dots,ij\}})$$

The proximity value p_{ij} is then defined as:

$$p_{ij} = \begin{cases} d(\Delta F_{ij}) & if\ B_i = B_j \\ \Delta B_{ij} + d(F_i + F_j) & otherwise \end{cases}$$

For each edge f_{ij} the proximity index p_{ij} is a measure of difference in floors ΔF_{ij} and geodetic distance between centroids of buildings ΔB_{ij}. Acknowledging that linear horizontal distance between buildings is experientially very different from vertical difference between floors, I nonetheless define the proximity value as a function of the two. A scale factor of d is introduced to adjust floor values such that they are comparable to distance values. However, for the scope of the present study, I do not attempt an empirical space syntax to better synthesize multiple distance metrics (as in Hillier 1984; Peponis 1998, 2007). Defining a better quantitative measure of functional distance is a promising avenue of further research.

b. Layer 2: *Co-authorship Index*

The publication value is summarized as a scalar index of scientific productivity. Each node corresponds to one author i and a link between two authors i and j exists if and only if the two authors have collaborated at least one time on the same paper. Links are weighted with a value representing the intensity of collaboration between the respective actors. Given that the purpose of the network analysis is to better understand spatial relationships – for example, co-location and frequency of contact between collaborators – I include the total number of co-authors, presuming that the spatial dynamics of collaboration between a team of two is different from the dynamics between a team of ten (Newman 2001b).

The publication edge weight (b_{ij}) accounts for the number of co-authored papers between faculty i and j and the number of total co-authors for those papers (MIT faculty and non-MIT faculty). As such, the weight of a link summarizes not only co-productivity between collaborators, but also the way in which those collaborations were carried out. Suppose that two authors i and j have collaborated n_{ij} times in n_{ij} different papers, then the weight w_{ij} is given by:

$$w_{ij} = \frac{n_{ij}}{\sum_{k=1}^{n_{ij}}(n_k^{c_{ij}} - 1)}$$

Where $n_k^{c_{ij}}$ is equal to the number of total co-authors on paper k.

c. Layer 3: *Co-invention Index*

Similar to co-publication, a weighted link is used to represent intensity of co-patenting activity.

A summary of the size and characteristic properties of the co-authorship and co-invention networks is shown below, and detailed network characteristics are presented in Figure A12, including: node degree distribution, average clustering coefficient, and characteristic path length. Notable features include the following: Average Number of Neighbors indicates that MIT authors who collaborate at least once with another MIT faculty member (and, as such, participate in the network) have an average of 5.47 MIT collaborators over the 10 year period, while collaborative inventors have an average of 2.51. Furthermore, the percentage of Shortest Paths, as well as the Average Betweenness Centrality, evidence differences in the community structure of the two networks (Section 6b). Finally, the Node Degree Distribution follows a power law for both networks, suggesting that they are scale-free, although the size of the patent network is too small to make claims about its scale-free properties (Barabási 1999; Goh 2001). Some conspicuous features are apparent in the visualizations below – for example, that the Department of Chemical Engineering tends to be very central to the paper network, with a high average degree, while Electrical Engineering & Computer Science is widely distributed, but with a low degree. Furthermore, some groups emerge, such as the Sloan School of Management and the Department of Economics.

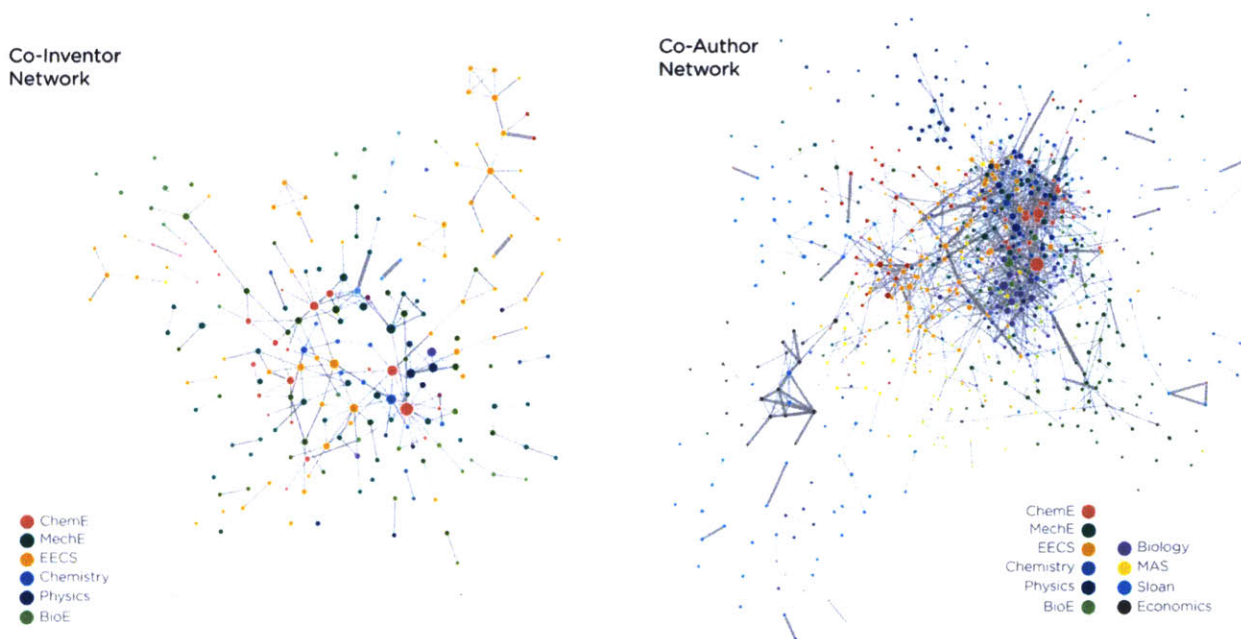|	Paper	Patent
Number of Nodes	710	215
Clustering Coefficient	0.233	0.234
Connected Components	19	22
Diameter	16	14
Shortest Paths	429820 (85%)	19526 (42%)
Characteristic Path Length	4.91	5.08
Average Number of Neighbors	5.47	2.51
Average Betweenness Centrality	0.016	0.131

Figure 9. Co-invention and co-authorship networks. Nodes represent faculty members, and node size is determined by its degree. A link indicates one or more collaborative papers or patents, and its thickness is the weight – a function of number of papers between the given authors, and the total number of authors on those papers. Nodes are colored according to department, and some notable departments are labeled.

Betweenness Centrality

The betweenness centrality of a node i is the number of shortest paths between other nodes that run through i. The measure was first proposed in 1977 by mathematician Linton Freeman (Freeman 1977) and is a defining characteristic of complex weighted networks (Barabási 1999; Goh 2001; Barrat 2004). Betweenness centrality can be interpreted as a measure of network resilience: it indicates how many geodesic paths will get longer when an arbitrary node is removed from the network. The betweenness centrality of a node i is given by the expression:

$$g(i) = \sum_{j \neq i \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$$

where σ_{jk} is the total number of shortest paths from node j to node k and $\sigma_{jk}(i)$ is the number of those paths that pass through i. The betweenness centrality can be averaged for all nodes in a graph to arrive at a characterization of the network. A high degree of betweenness indicates that certain faculty tend to operate as cross-pollinators within the network. I find an average betweenness centrality of the co-invention network to be 0.131 and the co-authorship network to be 0.016 This suggests that for patenting, it is more common for certain faculty to be crucial links, as the shortest path between other patenting faculty. Members of communities in the paper network are more densely interconnected, while the patent network has key nodes with a high Betweenness Centrality (Section 6b)

6b. Community detection and network topology

The fact of cross-pollination presumes distinct but interconnected groups. In network science, a cluster – or 'community' – is defined as a group of nodes with dense connections among its members, and fewer, weaker connections with external nodes. Our real-world network of collaboration at MIT is comprised of several such communities. Some are given – for example, faculty who are co-located in a building necessarily form a cluster in the spatial proximity network. Effectively identifying the community structure in the co-authorship and co-invention networks reveals groups of faculty who work together.

I test two categories of clustering method to find the community structure in the MIT collaboration networks – one based on agglomerative sorting and the other on divisive sorting. The former begins by classifying each individual node as a single community, and iteratively merges pairs of nodes with the highest similarity (as determined by a specified 'linkage method,' in our case, the edge weight w_{ij} as calculated in the overall network definition). The resulting classification can be intuitively represented as a dendrogram. The inverse is a divisive sorting algorithm, in which the edge with highest betweenness is removed, and global betweenness is recalculated, in an iterative 'cutting' process that is stopped at a certain community threshold.

What is that threshold? The definition of a community as 'more connected' with itself than the rest of the network is not sufficient to identify structure – it is necessary to specify a threshold of community membership. I compare a number of common sorting mechanisms (OSLOM, Infomap, Modularity, Louvain, Label, Hierarchical) to the random graph to identify the most effective community detection. I find that the OSLOM algorithm – Order Statistics Local Optimization Method, an algorithm that accounts for edge weights, overlapping communities, hierarchies and community dynamics (Lancichinetti 2011) – best structures the network into its constituent communities (Figure 10). Network topology can then be compared with the given building-level and department-level communities to understand the relationship between collaboration and co-location or co-affiliation.

Communities in
the Co-inventor
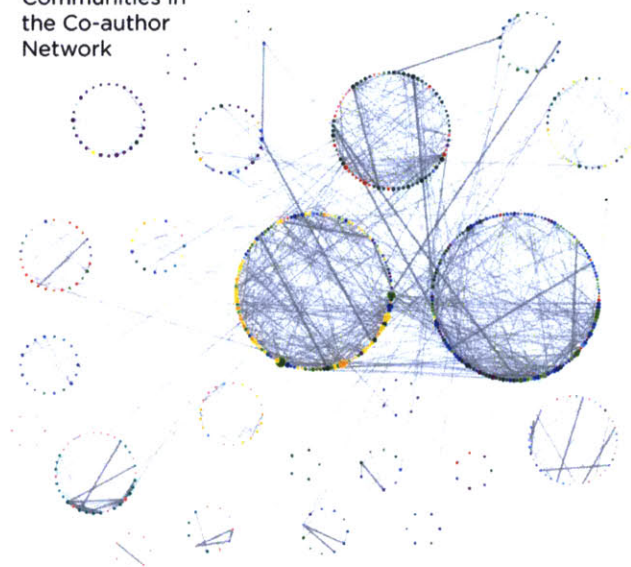Network

Communities in
the Co-author
Network

Figure 10. Community detection for the patent (left) and paper (right) networks. A community is a group of nodes that are more interconnected with each other than with the remainder of the network. To identify the community threshold, I apply the OSLOM algorithm and separate the resulting communities.

Table 7. Community diversity is a measure of the number of different buildings or departments represented in each group, normalized by community size. For the paper and patent networks, I find:

Diversity	Papers	Patents
Building	0.40	0.51
Department	0.29	0.70

Chemistry
ChemE
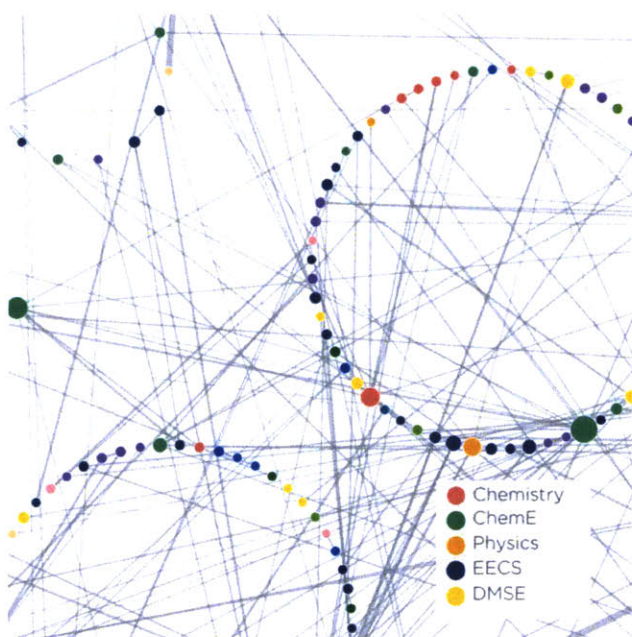Physics
EECS
DMSE

Sloan
Economics
MAS
Literature
EAPS
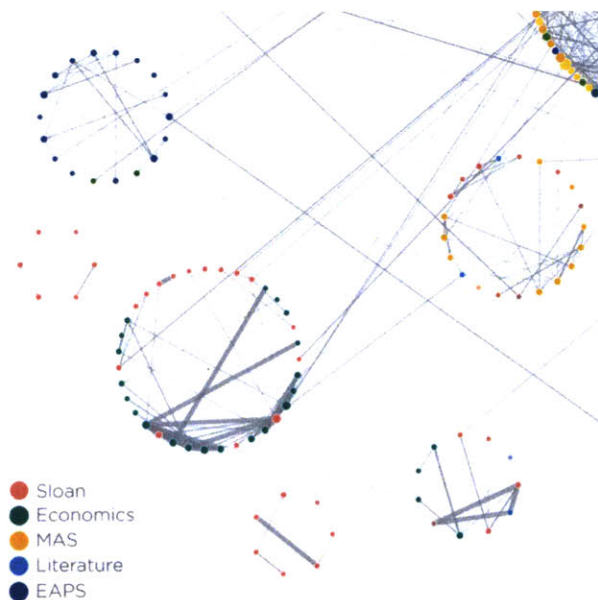
Figure 10a. Detail views of the co-invention and co-authorship networks. Communities in the former are more heterogeneous, comprising faculty from several different departments. Communities in the co-authorship network seem to be more topically defined, in this case, Earth Atmospheric & Planetary Sciences, or Sloan and Economics.

The Diversity of Collaborative Communities

Mapping a network that is clustered by communities of collaborative faculty is visually informative – for example, comparing a visualization in which nodes are colored by department with one in which nodes are colored by building. Moving beyond a general impression of diversity, I propose a simple quantitative measure of network topology based on assortment in community groups. The diversity of community structure in the network is given by the expression:

$$d_i = \frac{1}{n} \sum_{i=1} \frac{a_i}{p_i}$$

where n is the total number of communities, p_i is the number of members of community i and a_i is the number of affiliations in community i. A high d_i value indicates greater diversity in the community structure of a particular network – showing, for example, that communities in the co-authorship network tend to span across buildings more than they span across departments.

I find that this is, in fact, the case. The co-authorship network is characterized by a d value of 0.40 for buildings and 0.29 for departments, and the co-invention network by d values of 0.51 and 0.70. These attributes of network topology confirm the statistical results of Section 5a: co-authors tend to favor collaborators in the same department, while inventors collaborate within the building but across disciplinary boundaries. This points toward motivations for team composition – that co-inventors collaborate around projects, while co-authors collaborate within domains of scholarship.

6c. Proximity and Collaboration

It is evident that there are strong patterns of collaboration related to departments and buildings – but does simple proximity have an effect? Using the metric distance between collaborators, I test the hypothesis that proximity is correlated with co-authorship and co-invention (Section 6a). Across output types, the probability of a collaboration between two agents decays exponentially with their spatial proximity, to a remarkable degree of consistency. This is not dissimilar from, for example, the dissemination of dandelion seeds. One is more likely to find a seed close to the dandelion flower, and the likelihood of finding a seed decays exponentially as the distance increases. In the case of MIT faculty, one is more likely to find a collaborator close by. Co-authors have a higher tendency to collaborate at distance 0, suggesting co-authorship within a lab or department, while co-invention is more regularly spaced along the exponential curve. The probability $f(x)$ is given by:

$$f(x) = a\,e^{-bx}$$

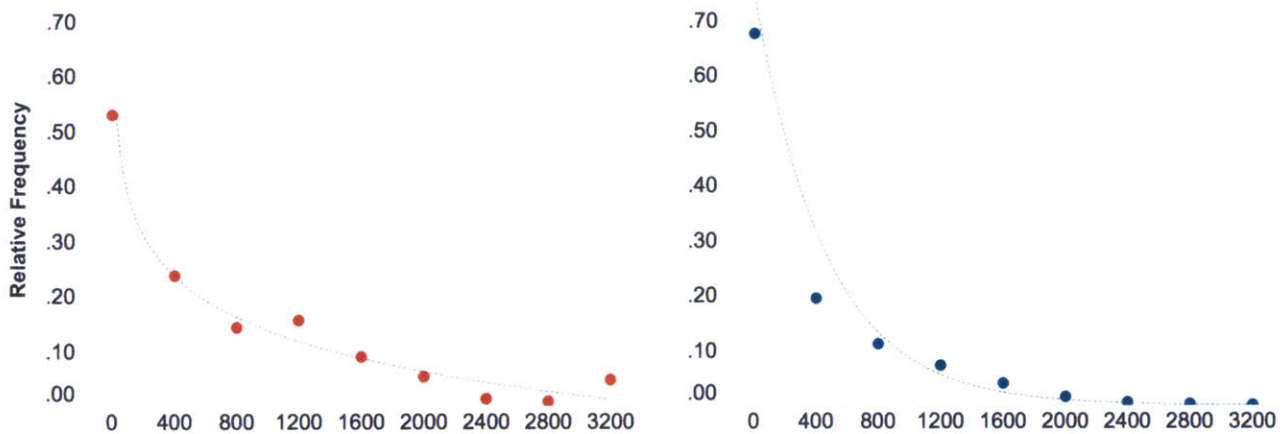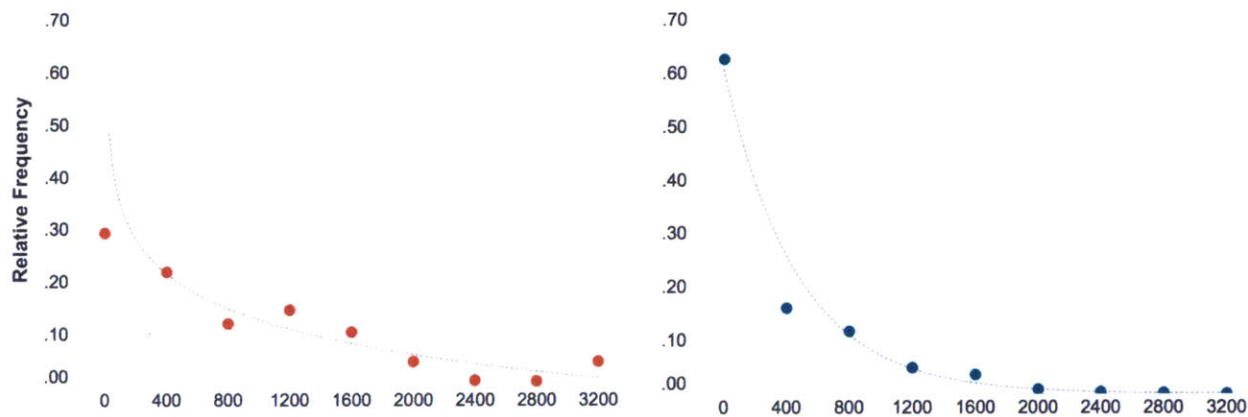|	Papers	Patents
	$a = 0.66$	$a = 0.72$
	$b = .0022$	$b = .0019$
	$R^2 = 0.98$	$R^2 = 0.99$

Figure 11. The relative frequency of collaborations between MIT faculty, plotted against their spatial distance on campus. As distance between two faculty members increases, the likelihood of their collaboration decreases according to a negative exponential function. The same pattern holds true for patents and papers.

Proximity and Cross-disciplinary Collaboration

One could imagine that the relationship between proximity and output as evaluated above – and particularly co-authorship at distance value 0 – is a function of co-location in the same building. That is, the relationship between proximity and collaboration may be skewed by tightly clustered and internally collaborative departments that are co-located. To control for this, I plot the same proximity function, specifically for pairs of collaborators from different departments. As with the total dataset, the decay of collaboration frequency over space – even in the specific case of cross-departmental collaboration – is modeled by a negative exponential function. For both papers and patents, the fitting curve has a strong statistical significance.

	Paper	Patent
R^2	0.98	0.86
p	<.0001	0.008

Figure 12. The relative frequency of collaborations between MIT faculty from different departments, plotted against their spatial distance on campus. As distance between two faculty members increases, the likelihood of their collaboration decreases according to a negative exponential function. The same pattern holds true for patents and papers.

As described in Section 5c., I examined the effect of overall central positioning on campus, and found no significant correlation with total output or with intra-building collaboration. That is, faculty members who are centrally located at MIT do not necessarily produce more papers or patents, on average, nor do they have a stronger tendency to collaborate with faculty in the same building versus outside the building. These results are not at odds with the finding that likelihood of collaboration decays exponentially over space. Being centrally located with respect to the entire community does not seem to have a strong effect on collaboration, but the proximity between specific faculty members does. That is, I observe a sharp contrast between the condition of *centrality* versus *proximity*.

6d. Distance limit or optimum

Distance Limit
Spatial proximity is crucial for initiating communication that may not otherwise occur between researchers in different intellectual domains, and for the development of a shared language that facilitates collaboration across disciplines (as in Latour 1986 among others). There is a strong relationship between the spatial proximity of faculty at MIT and their tendency to collaborate – a pattern that is consistent across papers and patents, for the entire MIT community, and specifically for cross-departmental collaborator pairs (Figures 11, 12). Given the persistence of the exponential model, I seek to further characterize the relationship between proximity and collaboration. Might there be a relevant distance *threshold* for faculty pairs – a meaningful limit beyond which regular in-person communication is too cumbersome? Or, is there an optimal median distance, as proposed in the proximity paradox of collaboration? This ideal distance would have a so-called 'goldilocks effect:' increased cognitive, organizational, social and institutional proximities benefit collaborative processes, but overwhelming or needless proximity suffocates innovative outcomes (Ben-Letaifa 2013; González-Ibáñez 2013).

Building on the concept of a time-based anthropological invariant in travel behavior proposed by Marchetti (1994) and the spatial cognate proposed by Kraut (2001), I hypothesize that there is a statistically significant distance limit for co-authorship, and an optimal median for co-invention. The proximity and collaboration curves (Section 6b) suggest this bias – co-author pairs are more densely distributed between zero and 500, suggesting collaborations internal to the building (and indeed, this is confirmed by the ratios in Section 5a.), while co-inventors are further spaced, even showing a rise near 1200-1600. As an initial test, I apply two mathematical tools to evaluate distance metrics: the e-fold distance and a cumulative density function for proximity.

e-Folding Distance
In fields as diverse as economics, chemistry, physics and ecology, observations of exponential growth or decay are described by a mathematical property called the *e-fold*. This is generally used to measure time, as an "e-folding lifetime" for example – the time it takes for a population to increase in number by a factor of e (2.71828). For a true exponential curve, this time span is constant for any given population size. I propose to use the same model for characterizing an "e-folding distance": the spatial interval across which the number of collaborations decreases by a factor of e^{-1}.

$$\frac{[A]_f}{[A]_i} = e^{-k(1/k)} = e^{-1} = 0.37$$

For the collaboration and proximity data, I find an e-fold value of 454 ft. for patents and 526 ft. for papers. Practically, this suggests that co-inventors who are an additional 454 ft. apart are 37% less likely to collaborate (a factor of e^{-1}). The 2e-fold values are 908 ft. for patents and 1052 ft for papers.

Cumulative Distance

To better evaluate the relationship between proximity and collaboration, and to identify a distance limit, I plot the cumulative density of distances for paper and patent collaborations. The resulting curves (Figure 13) show the sum of instances within a certain spatial distance, and percentile thresholds are listed below.
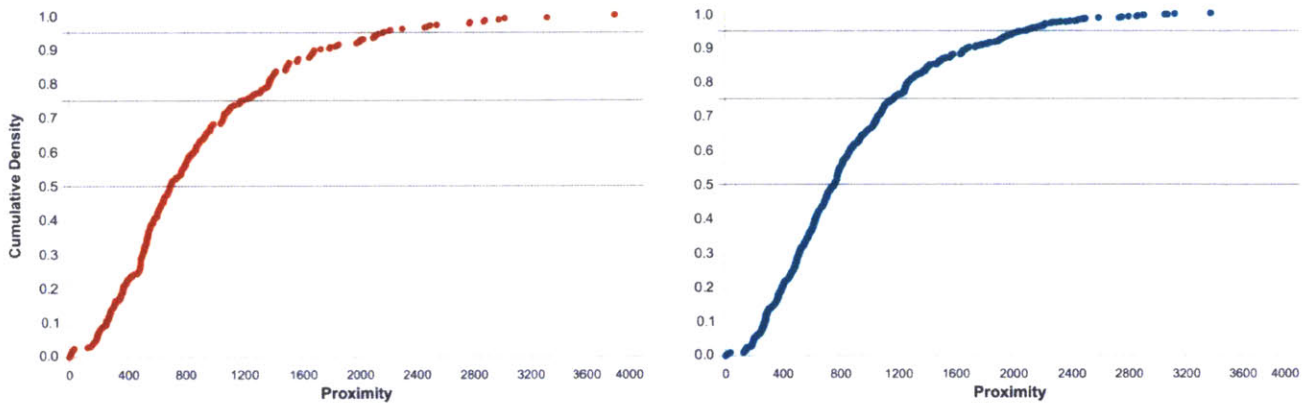
Percentile	Patent	Paper
95th	2165	2092
75th	1200	1170
50th	695	760

Figure 13. The cumulative density of proximities between MIT faculty collaborators. The 75th and 95th percentile are marked, as well as the average distance for both co-publishing and co-invention.

There appears to be an inflection point near the 75th percentile for both co-authorship and co-invention, corresponding to a proximity of ~1250. This can potentially be considered a meaningful distance for collaboration between faculty on MIT campus. I suggest further analysis to better determine the relevant distance limit.

7. Conclusion

Through statistical, geographic, and network analyses, I make a number of significant observations of the MIT research community, regarding its scholarly output and its socio-spatial collaboration dynamics. Between the years 2004 and 2014, MIT faculty produced over 40,000 papers and 2,300 patents, of which 6,414 and 454, respectively, involved more than one MIT faculty member. Broadly, an initial evaluation of this output confirms academic trends well recognized in the field of scientometrics. For example, there are observable differences in publication and patenting activity, a consistent temporal disparity between the two, and a strong divergence in departmental patterns. The fact of publishing a paper is different across fields, for example, and some areas of academia simply do not produce patents. Department-level analysis evidences these disparities clearly, even showing peaks of activity surrounding specific scientific breakthroughs – for example, the discovery of the Higgs Boson in particle physics.

In light of the literature on collaboration and co-location, I test several building-level attributes – including central positioning on campus, crowdedness, and heterogeneity of department affiliations – and find that their effects are not as pronounced as I anticipated. Intuitively, centrally positioned, densely populated and multi-disciplinary spaces would be active hotspots of collaboration, but these characteristics are not statistically correlated with productivity, using the measures applied in this study. Nevertheless, I observe that those buildings with the greatest raw number of departments are also outliers in productivity and collaboration. These outliers, then, can be understood as hotspots, despite that there is no global linear correlation.

Over the 10-year period of analysis, the MIT community as a whole published and patented with an increasing number of total collaborators on a given document. For both papers and patents, the rate of collaboration between MIT faculty has remained relatively constant, save for three notable exceptions: in 2009 and 2014, there was more co-invention among MIT faculty, and in 2012 there was more internal co-authorship.

A document-level analysis reveals very different patterns for paper vs. patent collaboration within departments and buildings. I calculate a metric for diversity of team composition on a given document, accounting for the number of collaborators, and the number of different buildings or departments that those collaborators represent. Overall, building diversity is higher than departmental diversity for both paper and patent teams, and co-inventor teams are much more heterogeneous than co-author teams, in both dimensions. The two output types exhibit diverging trends over time – patent teams becoming more homogeneous and paper teams the inverse. Patent team compositions are highly irregular over the observed time frame, and specific fluctuations may be linked to known factors. For example, the steep rise of intra-building collaboration in 2012 may be related to the inauguration of the Koch Institute for Integrative Cancer Research, a building intended to promote exactly that activity. This observation primes a number of questions: was it the fact of spatial instability that caused more intra-building collaboration? Or is it a result of the stated aim of the Koch Institute, to promote cross-disciplinary work? Is it a function of the specific faculty members themselves, access to specialized equipment, or intramural vs. extramural placement? The dynamics and effects of this shift are addressed in detail in my proposal for future research.

To further analyze collaborations within the MIT scholarly community, I structure a multi-layer network, with links representing a number of pairwise relationships between actors – co-authorship, co-invention and spatial proximity – and node attributes such as building and department affiliation. A mathematical analysis shows that the co-authors network is more densely interconnected than the patent network, and

that, conversely, inventors have a higher average *betweenness centrality*. That is, the patenting network is structured such that there are key members of ecosystem who function as cross-pollinators between distinct communities.

A deeper evaluation of network topology reveals the composition and interaction of these collaborative clusters. I first abstract the network into communities using an empirically defined algorithm (the OSLOM community detection algorithm, a method that accounts for edge weights, overlapping communities, hierarchies and community dynamics). Taking the resulting communities as a unit of aggregation, I quantify several features. In the publication network, communities are more strongly related to departments, whereas in the patent network, communities are more heterogeneous overall, and have a tendency to align with buildings. This suggests that co-inventors organize around ideas or projects, while co-authors organize around disciplinary areas – a finding that reconfirms the statistical results related to team composition. Overall, these findings point toward faculty members' practical motivations for working together: co-inventors collaborate around projects, benefitted by shared equipment and a breadth of expertise, while co-authors collaborate within domains of scholarship to advance the knowledge of a particular subject.

Finally, I test the effect of spatial proximity on collaboration. Regardless of co-location in a specific building, I find a persistent relationship between spatial proximity and the frequency of collaboration, well fit with an exponential decay model. Building on that observation, I suggest the possibility of a significant 'distance limit' for co-authors and an optimal 'goldilocks distance' between inventor pairs. Moving forward, the present analysis could be enhanced by more effectively isolating variables, conducting a time-based analysis, and addressing communication between collaborators.

7a. Future research

There are a number of improvements to this analytical structure, as well as promising avenues for further research and practical application. To these ends, I suggest a strategy to more effectively isolate variables (difference in difference analysis), to introduce a time variable in the network analysis (a dynamic systems approach), and to evaluate the processes of collaboration, in addition to its products (using digital communication among the community). Together, these can provide an understanding of academic dynamics over time, revealing the impact of spatial or institutional shifts on research at MIT.

Isolating Variables: Difference in Difference Analysis
The present study fails to isolate a number of variables, and as a result, is descriptive rather than explanatory. For example, any effects of department affiliation, physical location and output may be overshadowed by a funding grant that incentivizes collaboration. To address these concerns, I propose an analytical design based on a difference-in-difference analysis of specific shifts in physical configuration, drawing on precedent such as Catalini (2012) and Wineman (2014).

A natural experiment can be structured around the discrete spatial shift associated with the inauguration of the Koch Institute for Integrative Cancer Research (*Case A*) – a new funding resource, institutional entity and research facility, introduced in multiple phases. Similarly, the discrete institutional shift associated with the MIT Energy Initiative (*Case B*) provided funding and a cross-disciplinary linkage, but without a physical space.

Case A: Koch Institute
Following a $100 million grant from David H. Koch in October 2007, the MIT Center for Cancer Research was officially renamed to the David H. Koch Institute for Integrative Cancer Research. The first shift connected researchers who were formerly disparate in terms of affiliation, and provided funding for new research partnerships. In December 2010, a new building was inaugurated, bringing *some* of those researchers together into a shared lab facility. To date, researchers associated with the Koch Institute are divided between intramural and extramural faculty.

Case B: MIT Energy Initiative
In 2006, MIT Energy Initiative (MITei) was launched with the explicit goal of fostering cross-disciplinary and cross-departmental research related to energy. The initiative links a large number of faculty with a focus on energy research, and provides various financial and institutional support, but does not have a dedicated space.

Considering each discrete shift as a natural experiment, I propose to examine changes in the scholarly output of faculty members over time, as a continuum before and after the shift. Results of the present study suggest that this is a promising avenue of research – specifically through the observed patterns of team composition surrounding the inauguration of the Koch Institute. It may be impossible to determine causality, but I hypothesize that changes in office configuration will be associated with characteristic patterns in academic activity. In the case of the Koch Institute, for example, it will be possible to control for funding, as faculty were affiliated to the institute for several years before relocation to the building, and for space, because only a subset of the faculty were relocated to the new building in 2010. A time-based analysis of output and collaboration over the course of organizational transitions, using both MITei and the Koch Institute, may reveal the effect of spatial and institutional configurations on research.

Systems Approach: Time Element of Networks
Scientific collaboration can be considered as a dynamic, evolutionary system. Building on the analysis of community structure, I propose that the growth and change processes of the MIT collaboration networks – adding new nodes, shifting the connections between existing nodes – are expressed over time (Barabási 2002). In the same way as a multilayered network (co-authorship, co-invention and proximity) is structured through the rewiring of edges between a given set of nodes, a time-dependent network could be structured for temporal change in a single variable (rewiring edges between the node set according to publications from each year) and adding new nodes. A systematic investigation of changing network topology, including clustering coefficient, average degree, and scaling behavior, as well as the proposed community diversity metric (Section 6b.), could shed light on the complex evolution of scientific co-authorship at MIT.

Processes and Products: Digital Communication as Indicator of Collaboration
Finally, I propose an analysis of collaboration *processes* to contextualize and enrich the exploration of academic *products*. A significant portion of the contemporary team-based research process is conducted via telecommunication, primarily as email. I anticipate that characteristic patterns of email activity between faculty are related to collaborative authorship and invention. Furthermore, a strong body of literature suggests that these communication patterns are related to physical configuration and spatial shifts (Allen 1977, 1986; González-Ibáñez 2013; Kabo 2014, 2015; Gilson 2015 and others). For example, I anticipate an increase in contact between two researchers following a relocation event, and a higher likelihood of subsequent collaborative output.

Following a similar procedure to the proximity, co-authorship and co-invention networks, a new set of weighted links can be structured, representing *communication intensity* among faculty. For nodes i and j, a directed link between i and j represents the fact that i has sent at least one email to j. The value of the link weight will represent the intensity of communication from i to j, a function of both the number and the volume of messages sent. Proper normalization will be applied to account for the disparity in magnitude of these two quantities, similar to the co-authorship and co-invention link weight calculations.

7b. Impact

This empirical analysis of interrelated networks and academic output at MIT advances the broader knowledge of complex relationships between architecture, institutions and innovation. Socio-spatial network analysis is of practical interest to MIT's scientific community, to campus planners and administrators, and to the individual faculty members who are engaged in research activity. It is particularly salient in the context of new buildings, funding structures and initiatives that span the institute, as well as the growing entrepreneurial resonance of knowledge spillovers – research that exits the lab and is commercialized in the surrounding district.

Team science is increasingly multidisciplinary and institutionally atomized, and traditional academic structures such as departments with dedicated buildings may soon be anachronistic. I envision data-driven spatial and institutional organization strategies that actively promote and strengthen place-based collaboration. Campus planning could be managed in a more flexible way, such that researchers are dynamically re-positioned based on time-dependent activity, communication intensity, trends in broader fields of science, or random allocation. In the future, empirically grounded, real-time spatial and organizational principles could be extended to a variety of contexts that have an agenda of supporting collaborative work: an applied spatial science for innovation ecosystems.

Appendix A: A History of Innovation at MIT

MIT – and its surrounding innovation ecosystem of Kendall Square and Greater Boston – is a global hub of innovation and entrepreneurship. Faculty in the academic community produce cutting edge research, much of it disseminated and commercialized in corporations or entrepreneurial ventures, via patents and papers. I implement these observable vectors to understand collaborative research activity at MIT – but that analysis is given context by a portrait of the institute's history and its current practices of knowledge production and dissemination.

A New Institutional Model

During the early years of the institute, many characteristics emerged that would later become important determinants of MIT's success as a center for research. During the 1840s, William Barton Rogers – then a material scientist at University of Virginia – began working on a new concept for a pedagogical model inspired by European research universities, what he termed a *polytechnic* university. The polytechnic would be a radical merger of disciplines, people, ways of creating knowledge, and paths for disseminating knowledge: it was a hybrid education that incorporated varied fields of study (chemistry, metallurgy, architecture, engineering), and different modes of learning (lectures, laboratory work, practical exercises, and excursions). Rogers consolidated these ideas in a pamphlet titled "Objects and Plan of an Institute of Technology," in 1860, and – at the founding of what was then known as Boston Tech a year later – Rogers realized his idea of the polytechnic Institute.

> **Methods and Apparatus of Instruction**[1]
> The instruction in this department of the School will be given through the medium of —
> 1. Lectures and Familiar Expositions.
> 2. Oral and Written Examinations.
> 3. Practice in Physical and Chemical Manipulations.
> 4. Laboratory Training in Chemical Analysis, Metallurgy, and Industrial Chemistry.
> 5. Drawing and the Construction of Special Plans and Projects of Machines and Works of Engineering and Architecture.
> 6. Practical Exercises in Surveying, Levelling, Geodesy, and Nautical Astronomy.
> 7. Excursions for the Inspection and Study of Machines, Motors, Processes of Manufacture, Buildings, Works of Engineering, Geological Sections, Quarries, and Mines.

As a result of Rogers' pedagogical innovation, the culture of MIT was, from its genesis, one that prized applied problem solving. Rogers believed that practical knowledge is of inherent intellectual value, and furthermore that it can only be obtained through first-hand experience. The culture of MIT, inspired by the founder and shared by students, faculty, and partners, was characterized by social responsibility and an enthusiasm for tackling difficult challenges. Education was treated as a tool, "conducive to the progress of invention and the development of intelligent industry," (Rogers 1860). The new institution merged education, research, and practice – an approach crystallized in its motto, *mens et manus*.

Place and Urban Infrastructure

MIT was very much defined by its founding model, but the activity of the fledgling university was also facilitated by the urban context and physical infrastructure. Even at the time, the Boston region was a

---

[1] Rogers, William Barton. "Scope and Plan of the School of Industrial Science," 1864

locus of scientific activity – the American Academy of Arts and Sciences was founded in Cambridge in 1780. While a professor at University of Virginia, Rogers frequently traveled to Cambridge in the course of his affiliation with the academy, and was motivated by the region's intellectual quality and burgeoning industrial and economic climate.[2] These were the proto-conditions of an entrepreneurial ecosystem, rooted in academia and industry.

By the early years of the 20th century, competitive universities around the country were beginning to introduce research into their educational models and developing modern laboratories to support the activity. The institute was pressured to expand beyond what was feasible in Back Bay, and, in 1916, MIT moved across the river to Cambridge, with a stated aim to improve its infrastructure. The 'New Technology Campus' was sited in Cambridge, amidst factories and industrial space – not only a reflection of Rogers' vision for applied research, but also a location that offered tangible opportunities for expansion.

The architecture was designed, from the outset, as an innovative spatial strategy to promote multidisciplinary thinking, collaboration, and interconnection. The 1916 campus plan by architect William Welles Bosworth was conceived as a central hub (the Killian Court dome) with interconnected and extensible wings, designed to promote easy access between 'buildings.' The sprawling building-as-campus, anchored by an 'infinite corridor,' was novel not only in its continuity, but also its integration with the urban context. The prevailing university planning model of the time was as discrete and insular enclaves (see: the institute down the road). In contrast, MIT was spatially and institutionally indefinite, porous to the surrounding city and capable of expanding (Jarzombek 2004) – not only spatially, but soon, expanding its ties to industry.

Funding, Partnerships, and the Technology Plan of 1919
Following the university's move to Cambridge, then-president Maclaurin commissioned the Division of Industrial Cooperation and Research to explore strategies for augmenting the partnership model. The *Technology Plan of 1919* formalized a structure for academic-industrial partnerships that enabled the institute to dramatically improve its tenuous financial position. The first cohort of 200 member firms included some of the largest corporations of the time: General Electric, AT&T, US Steel and US Rubber (Ndiaye 2007). The great success of these collaborations over the following decades led to the founding of an Office of Corporate Relations, as well as the Industrial Liaison Program. "MIT for a long time... stood virtually alone as a university that embraced rather than shunned industry."[3] The groundbreaking 'Tech Plan' enabled MIT to forge strong links outside the academic world, and its success broadly legitimized the process of technology transfer and the pursuit of entrepreneurship surrounding academia.

Years later, during World War II, two simultaneous forces – a lack of funding and a national demand for innovative technology to support the war effort – caused the institute to dramatically increase defense research partnerships. MIT quickly became the largest research and development contractor with the US Government, under the leadership of Vannevar Bush (then head of the United States Office of Scientific Research and Development, as well as the vice president of MIT and the dean of MIT School of Engineering) (Zachary 1997). The pressures of meeting the innovation demands "made necessary the formation of new working coalitions... between these technologists and government officials. These changes were especially noteworthy at MIT," (Roberts 1991).

---

[2] Andrews, Murphy, Rosco. "William Barton Rogers: MIT's Visionary Founder," *MIT Libraries: Institute Archives, 2012.*
[3] Ridley, Matthew. "A Survey of New England: a concentration of talent," *The Economist.* August 8, 1987.

More importantly, those coalitions did not wane with the end of the war. MIT research expenditures according to primary funding shows US government-sponsored activity continuing vibrantly (Lecuyer 1992). Only much later, in response to protests against the Vietnam war, did MIT transition classified research off campus, to the new Lincoln Labs facilities. Following a precedent set by Vannevar Bush in 1940, MIT leaders have continued to serve as presidential advisors on matters of science, technology and research. In 1991, MIT established a Washington DC office to work actively in national science policy and research funding.

MIT Today: an Innovation Ecosystem
The path dependence of MIT's founding and history informs the policies and programs in place today and provides momentum for its position at the forefront of innovation. The institute maintains, at its core, a focus on research, and key elements specifically support knowledge production and dissemination. Over the course of the 20th century, and into the 21st, MIT has capitalized on the availability of adjacent space in Cambridge (much of it post-industrial), to establish world-class research facilities and a rich array of infrastructural resources, including an on-campus nuclear reactor, a pressurized wind tunnel, biomedical research facilities, and a hydrodynamics towing tank.

In a less concrete way, lab-based institutional organization enables cutting-edge research by bringing together people with varied skill sets and levels of academic achievement around applied challenges. It is also amenable to industrial sponsorship. Through the Industry Liaison Program (ILP), there are now 800 companies formally involved with MIT, working with these labs to develop research projects, license discoveries, and consult with researchers. Approximately 20% of total research funding ($128M) was sponsored directly by industry partners during the year 2014, and MIT is the nation's top university for industry-funded R&D.[1]

MIT Research Expenditure by Primary Sponsor
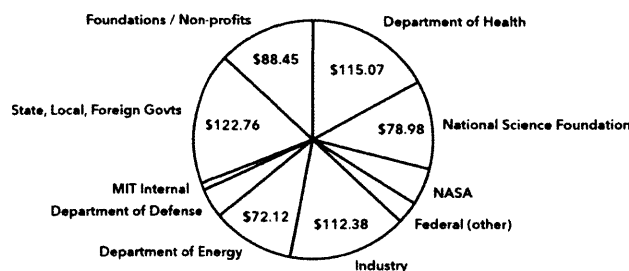
In addition to bringing funding, external partnerships enrich faculty engagement outside of academia. MIT has an intentionally porous boundary for knowledge application, supporting faculty who found businesses or engage in consulting through its Outside Professional Activities (OPA) policy and granting one day per week of flexible time. Furthermore, the Technology Licensing Office streamlines invention and commercialization by assisting with patents and intellectual property affiliated to MIT.

Adjacent to its focus on research and real-world application, MIT also supports entrepreneurship directly through education. Within the Sloan School of Management and across the entire institute, 48 entrepreneurship courses were offered during the 2013/14 year, and the institute maintains an overarching focus on cross-disciplinarity and team-based work. Not only do faculty members intellectually examine the processes of innovation (for example, in the Technological Innovation,

Entrepreneurship and Strategic Management group), but the university also brings on practitioners – the entrepreneurs themselves – to teach from experience. During the Summer and Independent Activities Period, a variety of non-credit courses, boot camps, accelerators and programs are offered to students and affiliates. For those who pursue entrepreneurial ventures, a large number of mentorship programs offer various forms of support, from early-stage ideation to scale-up, incorporation and technological development.

It is difficult to determine the extent to which entrepreneurial activity is directly attributable to these resources, but it is clear that the success of MIT's alumni is extraordinary. Over 30,000 active companies include an MIT alum among their founders. A quarter of graduates have participated in launching new companies, and 22% have worked as employees of early stage ventures. In terms of economic impact, firms founded by MIT alumni generate revenues of $1.9 trillion (equivalent to the GDP of the 10th largest global economy). The effect is also regional: over 40% of alumni stay in Massachusetts, and roughly 7000 alumni startups (31% of total) are located in the state.[4]

The local focus of entrepreneurship may be related to recent infrastructural expansion across Cambridge, and specifically in Kendall Square. The Kendall Square Redevelopment Plan, approved in 2013, called for re-zoning over one million square feet of land in support of the entrepreneurial ecosystem. The new Kendall Square district, currently under construction, will feature housing, retail, lab, academic and public space – in an area that remains fundamentally tied to the university. "The decision to promote commerce alongside research is a critical differentiator in MIT's success and competitiveness. One need only count the number of construction cranes, new restaurants, cinemas and museums in Boston to understand what a vibrant bridge between research and startups means overall city life."[5]

---

[4] Roberts, Murray, Kim. "Entrepreneurship and Innovation at MIT: Continuing Global Growth and Impact." 2015
[5] Enriquez, Juan. "Vassar and Main: the world's most innovative intersection," *Wired UK*, 02 Nov. 2015

Appendix B: Figures
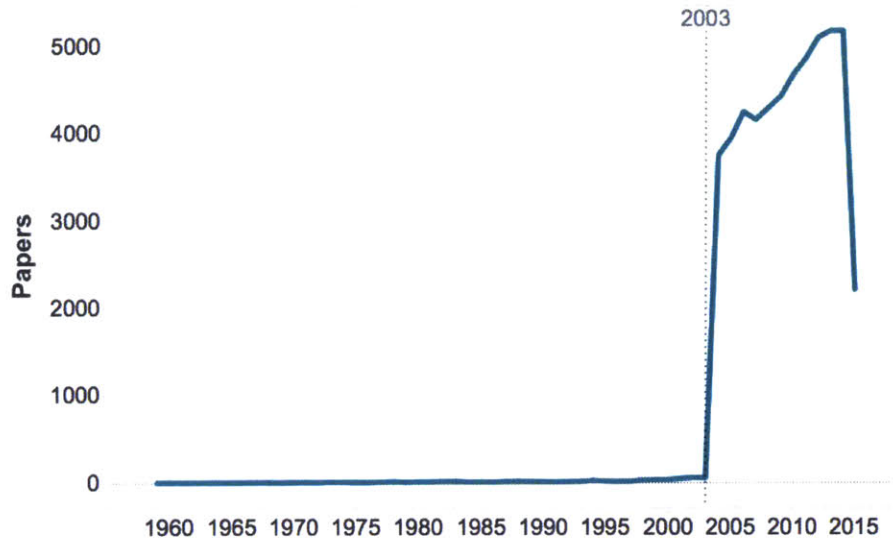
A1. The number of academic papers published by MIT authors (not exclusively faculty) per year, from 2004 to 2014. The dataset includes scholarly publications in peer-reviewed journals, with DOI identifiers. Given the shift in content management in 2003, the time period of analysis is limited to papers from the years 2004 to 2014.
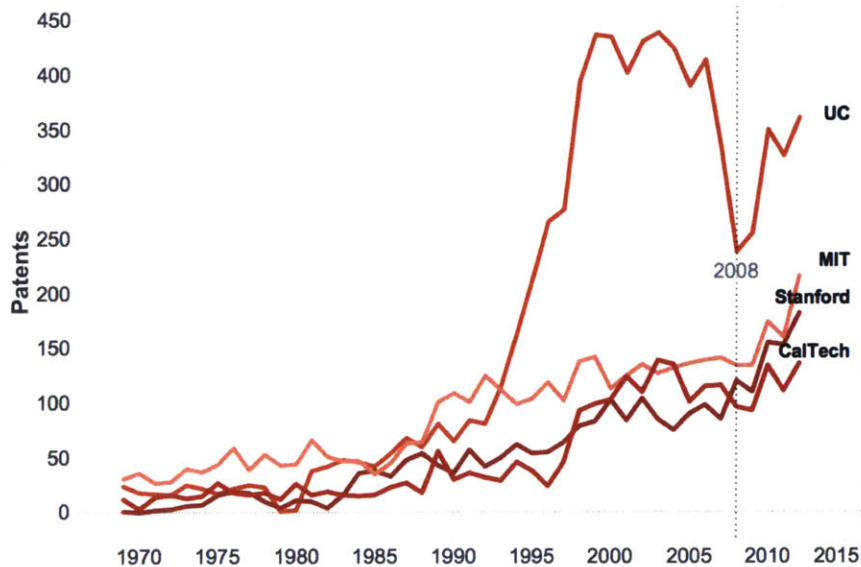

A2. MIT output to national university patenting trends (annual patent data from the US Patent and Trademark Office, "U.S. colleges and universities: utility patent grants calendar years 1969-2012"). For the top patenting research universities in the United States (MIT, Stanford, the University of California system, and Cal Tech)similar patterns are evident – specifically, a decrease around 2008.
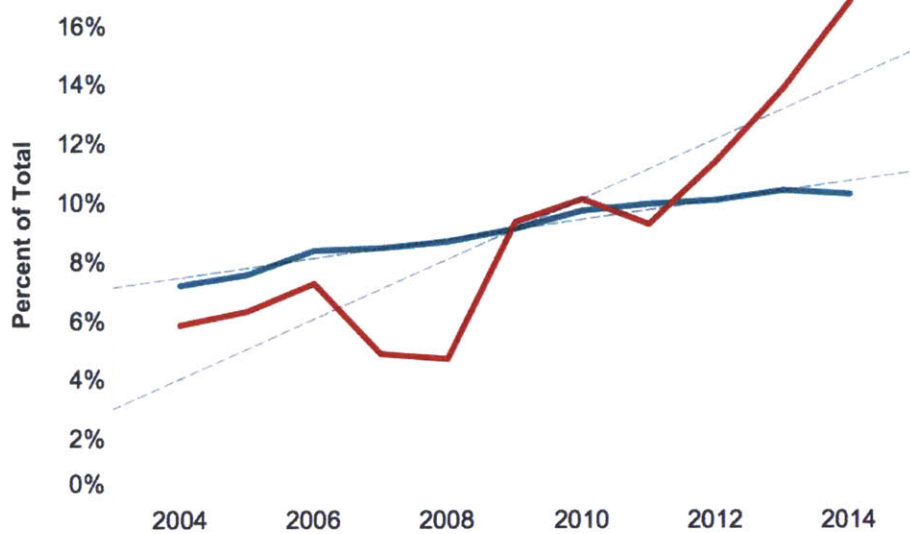
A3. The relative frequency of papers and patents per year (2004 – 2014), as a percentage of the total output during the time frame. Trend lines are calculated to show the rate of increase. Papers demonstrate an annual relative growth rate of approximately .3% and patents a rate of 1%
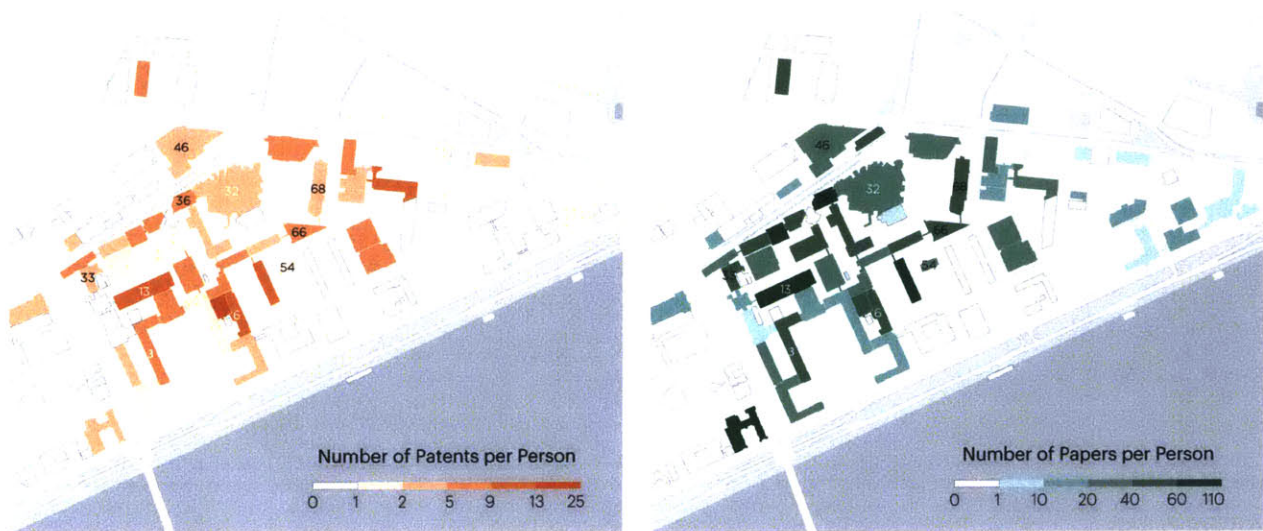

A4. Paper and Patent output per person, aggregated to the building level. Values are classified with a Jencks algorithm and mapped to campus buildings. Each building is labeled with its name (the facility key).
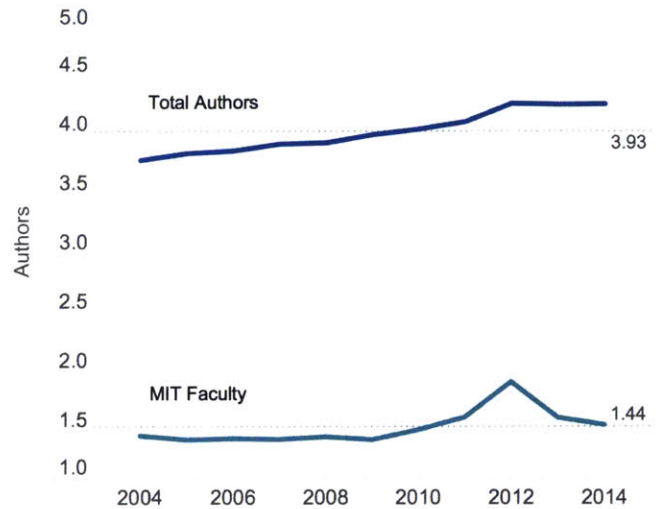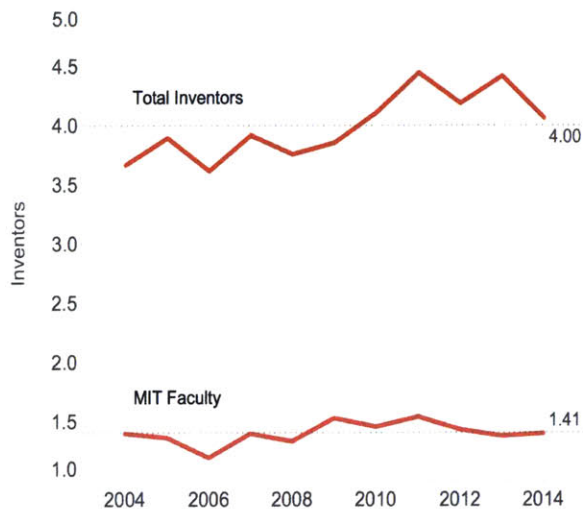
A5. The average number of total co-inventors per patent per year and the average number of total co-authors per paper per year, from 2004 to 2014. Colors differentiate collaborators who are MIT faculty and those who are not. There are an average of 4.00 total co-inventors and 1.45 MIT faculty co-inventors per patent over the time period, and a variable increase in total inventor numbers. There are an average of 3.95 total co-authors and 1.23 MIT faculty co-authors per paper and a linear increase in the total.
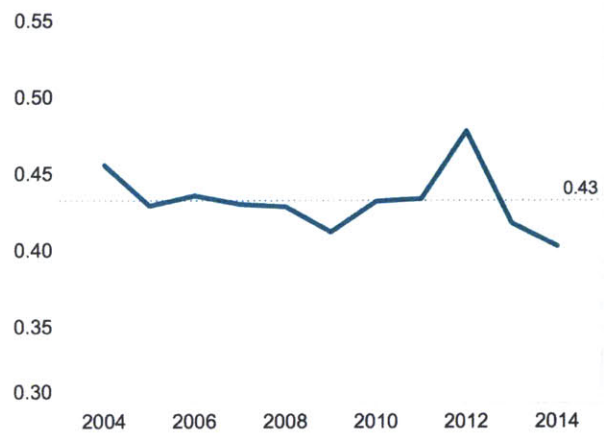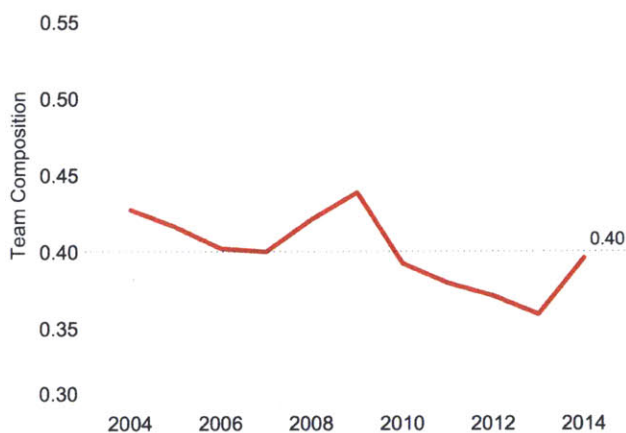
A6. The ratio of faculty to non-faculty collaborators for all MIT papers and patents per year. A value of 1 represents a document with only MIT faculty collaborators, and a value approaching 0 would indicate only non-MIT faculty collaborators (0 is impossible, as the dataset is based on MIT faculty output). Ratios for the entire community are aggregated by year, showing that collaboration ratios have decreased slightly over time. There are notable exceptions in 2009 for patents and 2012 for papers.
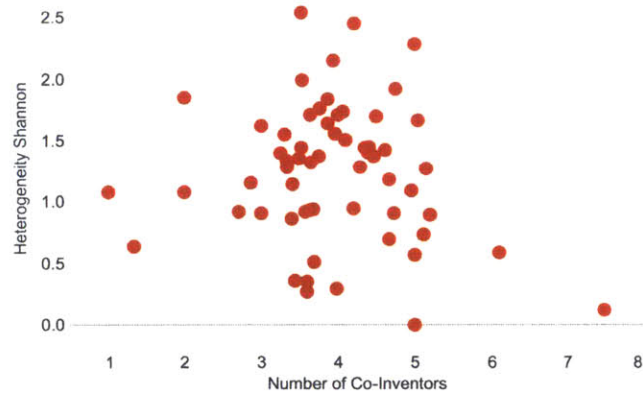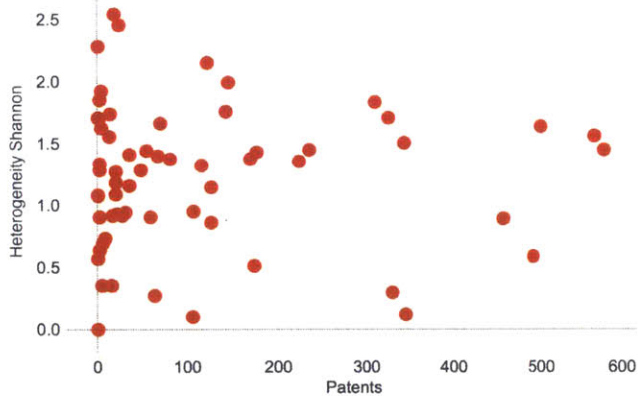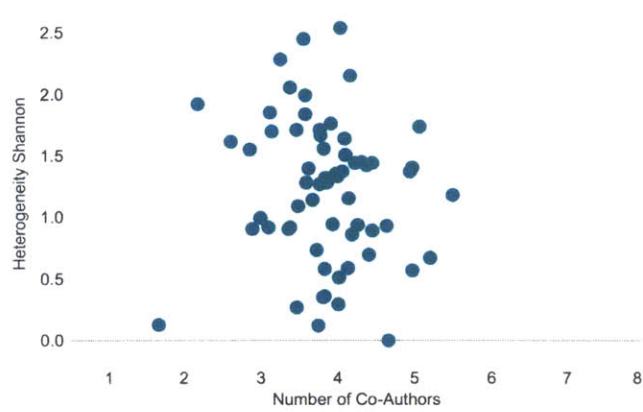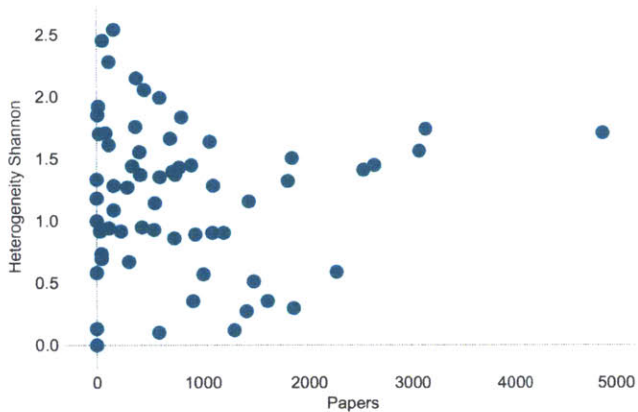
A9. Building heterogeneity, quantified using Shannon's Heterogeneity of Information, against total paper and patent output. Points represent individual buildings. There is no significant correlation between the variables.

A9. Building heterogeneity against total co-author and co-inventor numbers (aggregated to the building level). Points represent individual buildings. There is no significant correlation between the variables.
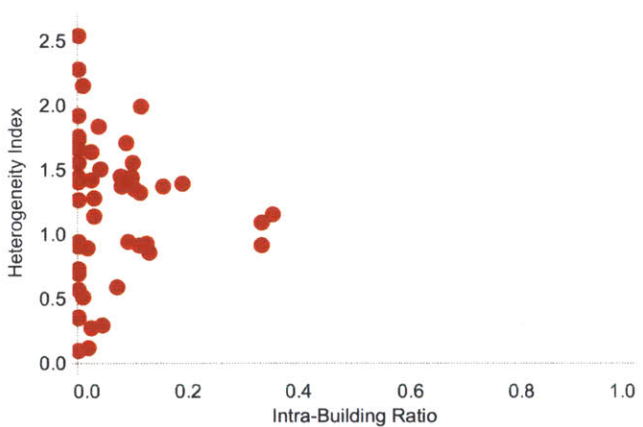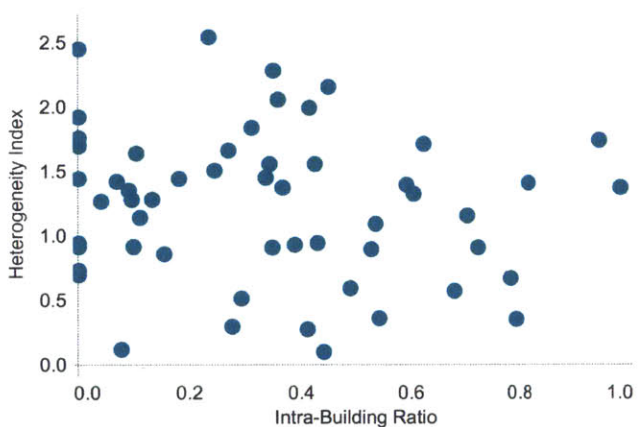
A9. Building heterogeneity against the ratio of internal to external collaborations, for both papers and patents. Points represent individual buildings. There is no significant correlation between the variables.
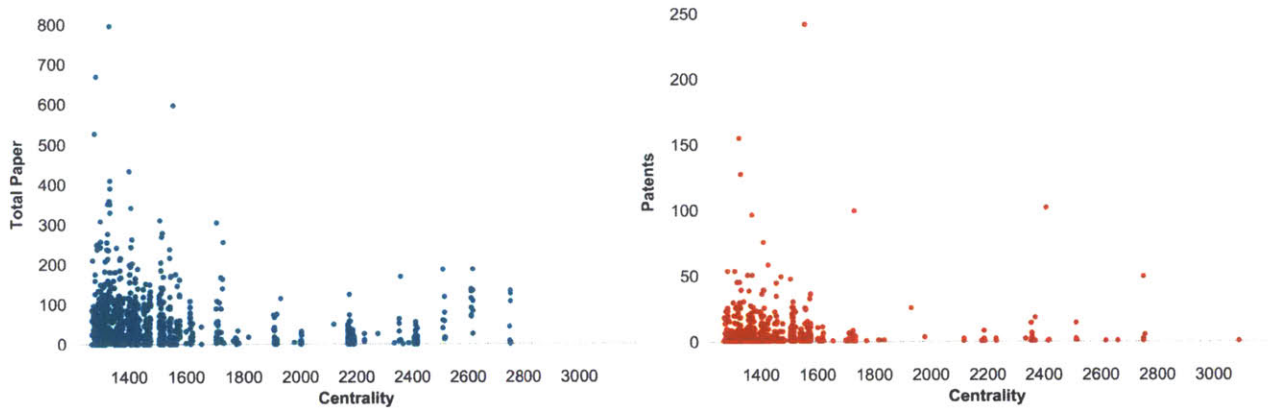
A10. Paper output per faculty, plotted against centrality, a measure of average distance to all other publishing faculty members. There is a weak correlation between higher centrality (lower average proximity) and paper output. The pattern is similar for patents.
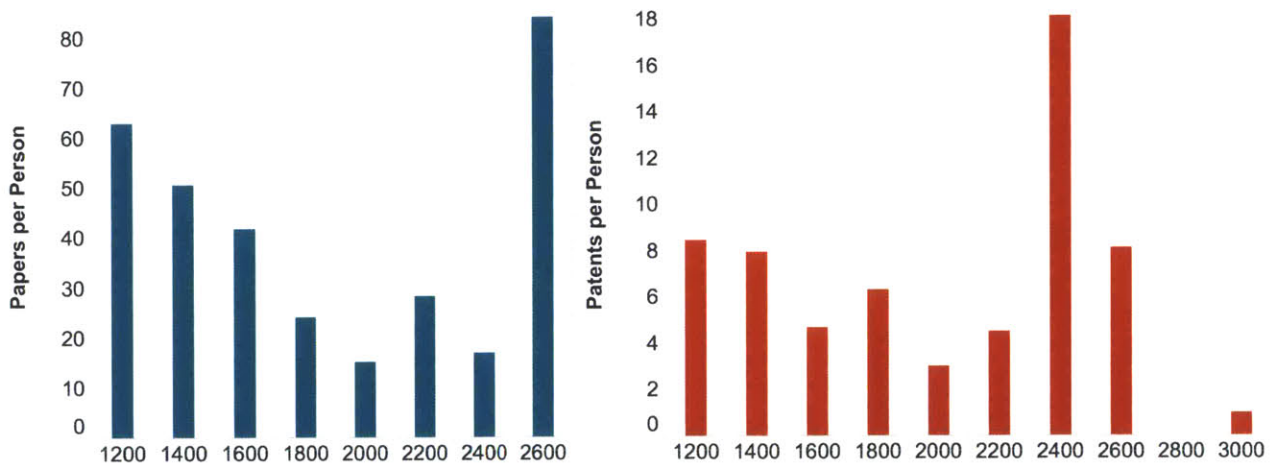
A11. Centrality bins are calculated for faculty, such that an average paper output can be determined for a given group of faculty with corresponding centrality. There is a similar pattern for papers and patents.
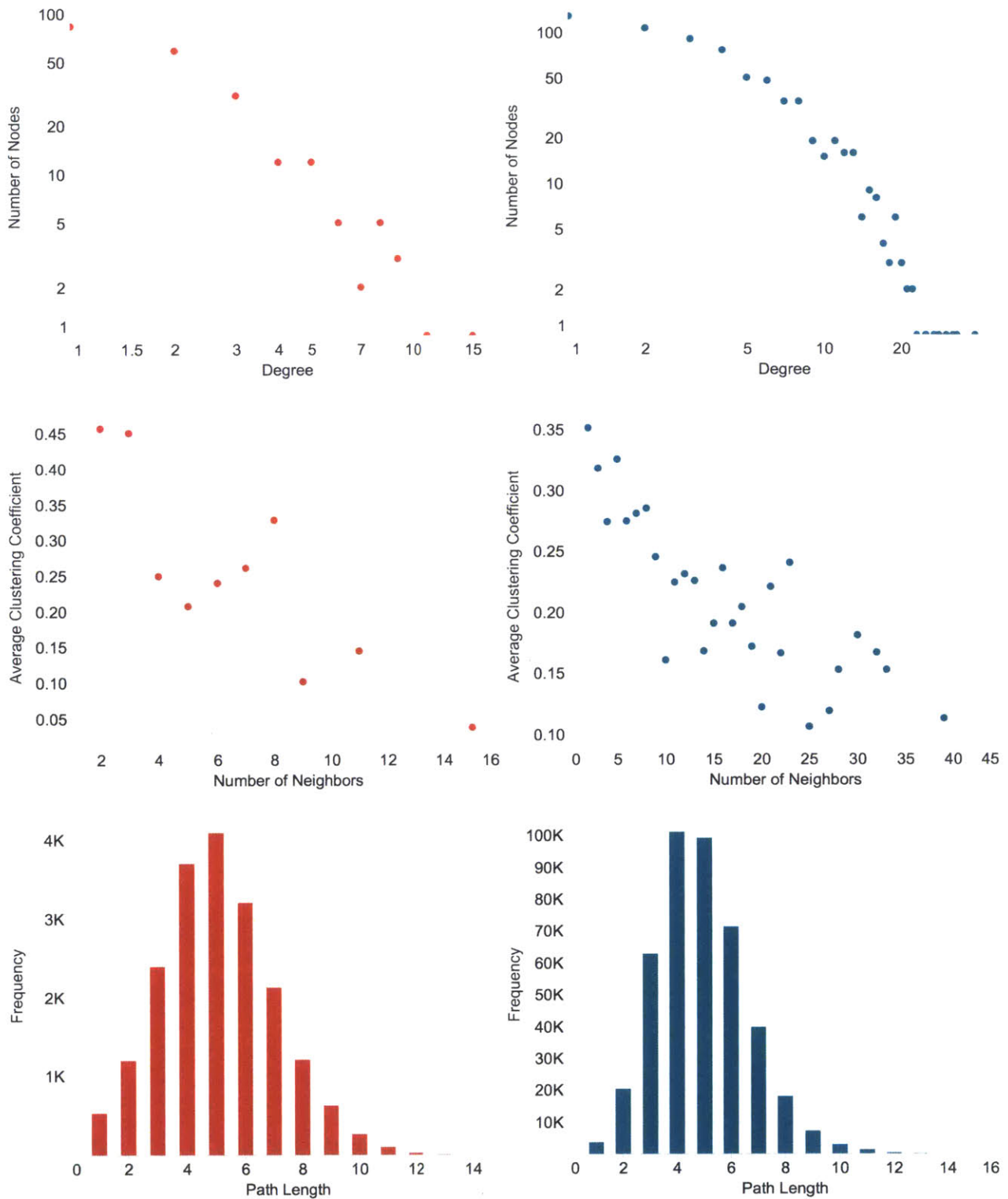
A12. Characteristics of the co-authorship and co-invention networks. These show: node degree distribution, average clustering coefficient, and characteristic path length. Node degree distribution is plotted on a log/log scale, demonstrating a power law that suggests scale-free network characterization.

Table A1 Department	Patent to Paper Ratio	Number of Total Authors	Papers Per Person	Papers	Average Annual Increase	Number of Total Inventors	Patents Per Person	Total Patents	Average Annual Increase
Aeronautics and Astronautics	.025	38	55.8	2121	3.85	15	3.7	55	0.78
Anthropology Program	.000	4	6.8	27	0.11				
Archaeology	.000	2	4.0	8	0.04				
Architecture	.014	20	9.1	181	0.96	3	1.0	3	0.00
Biological Engineering	.097	23	54.5	1253	7.74	22	5.4	119	0.53
Biology	.036	60	59.3	3557	14.52	33	3.8	125	0.34
Brain & Cognitive Sciences	.051	42	43.1	1812	5.85	18	5.2	93	1.75
Chemical Engineering	.132	35	80.2	2806	14.47	26	15.0	389	5.03
Chemistry	.096	34	79.1	2690	2.87	22	11.7	257	1.70
Civil and Environmental Engineering	.007	42	39.1	1641	10.77	11	1.2	13	0.14
Earth, Atmospheric & Planetary Sciences	.001	44	37.2	1636	8.00	1	1.0	1	--
Economics	.000	36	21.7	781	1.18	0	0	0	0
Electrical Engineering & Computer Science	.069	135	58.5	7893	13.95	90	6.2	558	6.25
Engineering Systems Division	.000	10	18.7	187	3.18	0	0	0	0
Foreign Languages & Literature Section	.000	2	1.5	3	0.00	0	0	0	0
Harvard-MIT Division of Health Sciences &Technology	.086	7	50.6	354	3.68	5	7.6	38	0.90
History Section	.000	7	1.4	10	0.11	0	0	0	0
Linguistics & Philosophy	.000	24	12.0	288	-0.12	0	0	0	0
Literature Section	.000	6	4.5	27	0.40	0	0	0	0
Materials Science & Engineering	.117	35	73.0	2555	7.76	28	10.6	296	1.03
Mathematics	.021	61	18.1	1106	3.91	5	4.4	22	-0.21
Mechanical Engineering	.098	79	48.1	3797	26.51	54	7.3	392	5.09
Music & Theater Arts Section	.000	1	2.0	2	0.00	0	0	0	0
Nuclear Science & Engineering	.010	22	26.1	575	3.51	4	1.5	6	0.03
Ocean Engineering	.000	10	9.5	95	5.00	0	0	0	0
Physics	.019	82	54.5	4471	3.85	6	14.5	87	0.25
Political Science	.000	25	5.6	140	0.11	0		0	0
Program in Media Arts and Sciences	.135	22	47.4	1043	7.72	17	8.9	151	2.20
Program in Science, Technology & Society	.000	13	8.1	105	0.11	0	0	0	0
Program in Writing & Humanistic Studies	.000	2	1.0	2	0.00	0	0	0	0
School of Engineering	.027	7	12.0	84	3.03	3	3.3	10	4.00
Sloan School of Management	.003	112	11.7	1306	4.43	3	1.3	4	0.00
Urban Studies & Planning	.010	23	9.5	219	-0.04	1	2.0	2	0.00

Table A2	Paper Entropy	Paper SD	Paper Variance	Paper Average	Patent Entropy	Patent SD	Patent Variance	Patent Average
Department								
Aeronautics & Astronautics	2.37	5.58	31.13	4.49	0.89	0.70	0.49	0.46
Anthropology Program	0.64	0.88	0.78	1.19	0.00	0.00	0.00	0.00
Archaeology	-	0.07	0.00	0.73	0.00	0.00	0.00	0.00
Architecture	1.22	1.04	1.09	1.12	0.34	0.42	0.17	0.13
Biological Engineering	1.32	4.05	16.39	1.85	1.82	5.01	25.14	2.79
Biology	2.68	5.39	29.01	5.69	1.39	2.74	7.50	1.39
Brain & Cognitive Sciences	2.36	4.07	16.59	4.35	0.81	1.35	1.82	0.61
Chemical Engineering	1.87	8.70	75.69	4.44	1.71	5.19	26.98	2.61
Chemistry	2.17	7.15	51.19	5.22	1.82	17.03	290.07	4.44
Civil & Environmental Engineering	2.14	4.52	20.47	3.65	0.50	0.47	0.22	0.23
Earth, Atmospheric & Planetary Sciences	2.17	3.49	12.16	3.34	0.24	0.31	0.10	0.07
Economics	1.86	2.24	5.04	2.40	0.00	0.00	0.00	0.00
Electrical Engineering & Computer Science	2.72	5.93	35.19	5.76	1.47	2.96	8.79	1.40
Engineering Systems Division	1.28	2.78	7.74	2.64	0.00	0.00	0.00	0.00
Foreign Languages & Literature Section	0.00	0.25	0.06	0.75	0.00	0.00	0.00	0.00
Harvard-MIT Division of Health Sciences and Technology	1.28	2.58	6.66	2.50	1.49	1.24	1.53	1.44
History Section	0.64	0.33	0.11	0.77	0.00	0.00	0.00	0.00
Linguistics & Philosophy	1.03	2.34	5.47	1.54	0.00	0.00	0.00	0.00
Literature Section	0.96	0.60	0.36	1.24	0.00	0.00	0.00	0.00
Materials Science & Engineering	2.28	7.08	50.09	5.13	1.91	3.23	10.44	2.28
Mathematics	1.64	2.59	6.70	1.94	0.19	0.41	0.17	0.10
Mechanical Engineering	2.04	5.19	26.99	3.10	1.82	3.86	14.87	2.07
Music & Theater Arts Section	0.69	0.30	0.09	0.70	0.00	0.00	0.00	0.00
Nuclear Science & Engineering	1.86	2.74	7.52	2.33	0.77	0.68	0.46	0.38
Ocean Engineering	0.00	0.06	0.00	0.03	0.45	1.05	1.10	0.66
Physics	2.75	12.20	148.78	7.59	0.47	2.20	4.82	0.49
Political Science	0.50	0.38	0.14	0.80	0.00	0.02	0.00	0.00
Program in Media Arts & Sciences	2.04	4.74	22.45	3.77	1.45	1.41	1.99	1.44
Program in Science, Technology & Society	1.07	0.94	0.88	1.10	0.00	0.00	0.00	0.00
Writing & Humanistic Studies	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
School of Engineering	0.80	18.43	339.60	7.88	0.41	0.34	0.12	0.19
Sloan School of Management	1.44	1.23	1.51	1.37	0.26	1.39	1.92	0.21
Urban Studies & Planning	1.14	1.58	2.51	1.26	0.52	1.78	3.17	0.45

Table A1. Patent and paper output by department, for the top six departments in each category. *Patent to Paper Ratio* represents the number of patents from a department divided by the number of papers, such that 0 represents a department without patents and 1 a department without papers.

Table A2. The unevenness of output per faculty member in each department. In some departments, a relatively small number of people generate the majority of the output, while in other departments, output is evenly distributed among faculty. *Average Annual Output Per Person* accounts for the number of years that a faculty member was affiliated with MIT (to control for the inherent bias toward those who have had a longer affiliation).

Works Cited

Allen, Thomas J. 1977. *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information within the R&D Organization.* Cambridge, Mass: MIT Press.

Allen, Thomas J. 1986. "Organizational Structure, Information Technology, and R&D Productivity." *IEEE Transactions on Engineering Management* EM-33 (4): 212–17.

Balland, Pierre-Alexandre, Ron Boschma, and Koen Frenken. 2015. "Proximity and Innovation: From Statics to Dynamics." *Regional Studies* 49 (6): 907–20.

Barabási, A.L. 1999. "Emergence of Scaling in Random Networks." *Science* 286 (5439): 509–12.

Barabási, A.L, H Jeong, Z Néda, E Ravasz, A Schubert, and T Vicsek. 2002. "Evolution of the Social Network of Scientific Collaborations." *Physica A: Statistical Mechanics and Its Applications* 311 (3-4): 590–614.

Barrat, A., M. Barthelemy, R. Pastor-Satorras, and A. Vespignani. 2004. "The Architecture of Complex Weighted Networks." *Proceedings of the National Academy of Sciences* 101 (11): 3747–52.

Baym, N. K., Y. B. Zhang, and M.-C. Lin. 2004. "Social Interactions Across Media: Interpersonal Communication on the Internet, Telephone and Face-to-Face." *New Media & Society* 6 (3): 299–318.

Ben-Letaifa, Soumaya, and Yves Rabeau. 2013. "Too close to Collaborate? How Geographic Proximity Could Impede Entrepreneurship and Innovation." *Journal of Business Research* 66 (10): 2071–78.

Bettencourt, Luís, M. A., J. Lobo, D. Helbing, C. Kuhnert, and G. B. West. 2007. "Growth, Innovation, Scaling, and the Pace of Life in Cities." *Proceedings of the National Academy of Sciences* 104 (17): 7301–6.

Bettencourt, Luís M. A., José Lobo, Deborah Strumsky, and Geoffrey B. West. 2010. "Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities." Edited by Juan A. Añel. *PLoS ONE* 5 (11): e13541.

Bikard, Michaël, Fiona Murray, and Joshua S. Gans. 2015. "Exploring Trade-Offs in the Organization of Scientific Work: Collaboration and Scientific Reward." *Management Science* 61 (7): 1473–95.

Björk, Bo-Christer, and David Solomon. 2013. "The Publishing Delay in Scholarly Peer-Reviewed Journals." *Journal of Informetrics* 7 (4): 914–23.

Borner, K., J. T. Maru, and R. L. Goldstone. 2004. "The Simultaneous Evolution of Author and Paper Networks." *Proceedings of the National Academy of Sciences* 101 (Supplement 1): 5266–73.

Bornmann, Lutz, and Rüediger Mutz. 2014. "Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References." *Journal of the Association for Information Science and Technology.*

Boschma, Ron. 2005. "Proximity and Innovation: A Critical Assessment." *Regional Studies* 39 (1): 61–74.

Boschma, Ron., and Ron Martin, eds. 2010. *The Handbook of Evolutionary Economic Geography.* Cheltenham, UK ; Northampton, MA: Edward Elgar.

Breschi, S., F. Lissoni, and F. Montobbio. 2007. "The Scientific Productivity of Academic Inventors: New Evidence from Italian Data." *Economics of Innovation and New Technology* 16 (2): 101–18.

Burt, Ronald S. 2004. "Structural Holes and Good Ideas." *American Journal of Sociology* 110 (2): 349–99.

Catalini, Christian. 2012. "Microgeography and the Direction of Inventive Activity." *SSRN Electronic Journal.*

Crouch, Dennis. 2011. "Person(s) Skilled in the Art: Should the Now Established Model of Team-Based Inventing Impact the Obviousness Analysis?" http://patentlyo.com/patent/2011/05/persons-skilled-in-the-art-should-the-now-established-model-of-team-based-inventing-impact-the-obviousness-analysis.html.

Delgado, Mercedes, Michael Porter, and Scott Stern. 2012. "Clusters, Convergence, and Economic Performance." w18250. Cambridge, MA: National Bureau of Economic Research.

Fazio, Catherine, Jorge Guzman, Fiona Murray, and Scott Stern. 2016. "A New View of the Skew: A Quantitative Assessment of the Quality of American Entrepreneurship." The Laboratory for Innovation Science & Policy: MIT.

Finholt, T. A., and G. M. Olson. 1997. "From Laboratories to Collaboratories: A New Organizational Form for Scientific Collaboration." *Psychological Science* 8 (1): 28–36.

Freeman, Linton C. 1977. "A Set of Measures of Centrality Based on Betweenness." *Sociometry* 40 (1): 35.

Gilson, L. L., M. T. Maynard, N. C. Jones Young, M. Vartiainen, and M. Hakonen. 2015. "Virtual Teams Research: 10 Years, 10 Themes, and 10 Opportunities." *Journal of Management* 41 (5): 1313–37.

Girvan, M., and M. E. J. Newman. 2002. "Community Structure in Social and Biological Networks." *Proceedings of the National Academy of Sciences* 99 (12): 7821–26.

Glaeser, Edward L., William R. Kerr, and Giacomo A.M. Ponzetto. 2010. "Clusters of Entrepreneurship." *Journal of Urban Economics* 67 (1): 150–68.

Goh, K.-I., B. Kahng, and D. Kim. 2001. "Universal Behavior of Load Distribution in Scale-Free Networks." *Physical Review Letters* 87 (27).

González-Ibáñez, Roberto, Muge Haseki, and Chirag Shah. 2013. "Let's Search Together, but Not Too Close! An Analysis of Communication and Performance in Collaborative Information Seeking." *Information Processing & Management* 49 (5): 1165–79.

Guimerà, R. 2005. "Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance." *Science* 308 (5722): 697–702.

Guzman, Jorge, and Scott Stern. 2015a. "Nowcasting and Placecasting Entrepreneurial Quality and Performance." w20954. Cambridge, MA: National Bureau of Economic Research.

Guzman, Jorge, and Scott Stern. 2015b. "Where Is Silicon Valley?" *Science* 347 (6222): 606–9.

Hall, Bronwyn H., and Nathan Rosenberg, eds. 2010. *Handbook of the Economics of Innovation. Vol. 2: [...]*. 1. ed. Handbooks in Economics 2. Amsterdam: North Holland.

Hillier, Bill, and Julienne Hanson. 1984. *The Social Logic of Space*. Cambridge: Cambridge University Press.

Hillier, Bill, and Alan Penn. 1991. "Visible Colleges: Structure and Randomness in the Place of Discovery." *Science in Context* 4 (01).

Hinds, Pamela, and Sara Kiesler, eds. 2002. *Distributed Work*. Cambridge, Mass: MIT Press.

Hsiehchen, D., M. Espinoza, and A. Hsieh. 2015. "Multinational Teams and Diseconomies of Scale in Collaborative Research." *Science Advances* 1 (8): e1500211–e1500211.

Huang, Mu-Hsuan, Huei-Ru Dong, and Dar-Zen Chen. 2012. "Globalization of Collaborative Creativity through Cross-Border Patent Activities." *Journal of Informetrics* 6 (2): 226–36.

Jensen, Richard, Jerry Thursby, and Marie Thursby. 2003. "The Disclosure and Licensing of University Inventions." w9734. Cambridge, MA: National Bureau of Economic Research.

Jones, Benjamin F., Stefan Wuchty, and Brian Uzzi. 2008. "Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science." *Science*. 322 (5905): 1259–62.

Kabo, Felichism W., Natalie Cotton-Nessler, Yongha Hwang, Margaret C. Levenstein, and Jason Owen-Smith. 2014. "Proximity Effects on the Dynamics and Outcomes of Scientific Collaborations." *Research Policy* 43 (9): 1469–85.

Kabo, F., Y. Hwang, M. Levenstein, and J. Owen-Smith. 2015. "Shared Paths to the Lab: A Sociospatial Network Analysis of Collaboration." *Environment and Behavior* 47 (1): 57–84.

Knoben, J., and L.A.G. Oerlemans. 2006. "Proximity and Inter-Organizational Collaboration: A Literature Review." *International Journal of Management Reviews* 8 (2): 71–89.

Knorr-Cetina, K. 1981. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science.* Pergamon International Library of Science, Technology, Engineering, and Social Studies. Oxford; New York: Pergamon Press.

Kraut, Robert, Carmen Egido, and Jolene Galegher. 1988. "Patterns of Contact and Communication in Scientific Research Collaboration." In 1–12. ACM Press.

Lancichinetti, Andrea, Filippo Radicchi, José J. Ramasco, and Santo Fortunato. 2011. "Finding Statistically Significant Communities in Networks." Edited by Eshel Ben-Jacob. *PLoS ONE* 6 (4): e18961.

Larson, James R., Caryn Christensen, Ann S. Abbott, and Timothy M. Franz. 1996. "Diagnosing Groups: Charting the Flow of Information in Medical Decision-Making Teams." *Journal of Personality and Social Psychology* 71 (2): 315–30.

Latour, Bruno. 2003. *Science in Action: How to Follow Scientists and Engineers through Society.* 11. print. Cambridge, MA: Harvard University Press.

Latour, Bruno, and Steve Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts.* Princeton, N.J: Princeton University Press.

Link, Albert N., John T. Scott, and Donald S. Siegel. 2003. "The Economics of Intellectual Property at Universities: An Overview of the Special Issue." *International Journal of Industrial Organization* 21 (9): 1217–25.

Marchetti, C. 1994. "Anthropological Invariants in Travel Behavior." *Technological Forecasting and Social Change* 47 (1): 75–88.

Milojević, S. 2014. "Principles of Scientific Research Team Formation and Evolution." *Proceedings of the National Academy of Sciences* 111 (11): 3984–89.

Miron-Spektor, Ella, Francesca Gino, and Linda Argote. 2011. "Paradoxical Frames and Creative Sparks: Enhancing Individual Creativity through Conflict and Integration." *Organizational Behavior and Human Decision Processes* 116 (2): 229–40.

Monjon, Stéphanie, and Patrick Waelbroeck. 2003. "Assessing Spillovers from Universities to Firms: Evidence from French Firm-Level Data." *International Journal of Industrial Organization* 21 (9): 1255–70.

Murray, Fiona, Edward B. Roberts, and J. Daniel Kim. 2015. "Entrepreneurship and Innovation at MIT: Continuing Global Growth and Impact." MIT Innovation Initiative.

Murray, Fiona, and Scott Stern. 2007. "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge?" *Journal of Economic Behavior & Organization* 63 (4): 648–87.

Newman, M. E. J. 2001a. "The Structure of Scientific Collaboration Networks." *Proceedings of the National Academy of Sciences* 98 (2): 404–9.

Newman, M. E. J. 2001b. "Scientific Collaboration Networks. I. Network Construction and Fundamental Results." *Physical Review E* 64 (1).

Newman, M. E. J. 2004. "Coauthorship Networks and Patterns of Scientific Collaboration." *Proceedings of the National Academy of Sciences* 101 (Supplement 1): 5200–5205.

Oerlemans, Leon, and Marius Meeus. 2005. "Do Organizational and Spatial Proximity Impact on Firm Performance?" *Regional Studies* 39 (1): 89–104.

Owen-Smith, Jason, and Walter W. Powell. 2004. "Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community." *Organization Science* 15 (1): 5–21.

Penn, A, J Desyllas, and L Vaughan. 1999. "The Space of Innovation: Interaction and Communication in the Work Environment." *Environment and Planning B: Planning and Design* 26 (2): 193–218.

Peponis, J., S. Bafna, R. Bajaj, J. Bromberg, C. Congdon, M. Rashid, S. Warmels, Yan Zhang, and C. Zimring. 2007. "Designing Space to Support Knowledge Work." *Environment and Behavior* 39 (6): 815–40.

Peponis, J, J Wineman, S Bafna, M Rashid, and S H Kim. 1998. "On the Generation of Linear Representations of Spatial Configuration." *Environment and Planning B: Planning and Design* 25 (4): 559–76.

Porter, Alan L., and Ismael Rafols. 2009. "Is Science Becoming More Interdisciplinary? Measuring and Mapping Six Research Fields over Time." *Scientometrics* 81 (3): 719–45.

Porter, M. E. 2000. "Location, Competition, and Economic Development: Local Clusters in a Global Economy." *Economic Development Quarterly* 14 (1): 15–34.

Porter, Michael E. 1998. *The Competitive Advantage of Nations: With a New Introduction.* New York: Free Press.

Porter, Michael, and Scott Stern. 2001. "Innovation: Location Matters." *MIT Sloan Management Review* 42 (4): 28–36.

Rallet, Alain, and André Torre. 2009. "Temporary Geographical Proximity for Business and Work Coordination : When, How and Where?" In 22. New York.

Rosenberg, Nathan, and Richard R. Nelson. 1994. "American Universities and Technical Advance in Industry." *Research Policy* 23 (3): 323–48.

Rothwell, Jonathan, José Lobo, Deborah Strumsky, and Mark Muro. 2013. "Patenting Prosperity: Invention and Economic Performance in the United States and Its Metropolitan Areas." Brookings Institute.

Shah, C., and R. González-Ibáñez. 2012. "Spatial Context in Collaborative Information Seeking." *Journal of Information Science* 38 (4): 333–49.

Stuart, Toby E., and Christopher Liu. 2010. "Boundary Spanning in a For-Profit Research Lab: An Exploration of the Interface Between Commerce and Academe." *SSRN Electronic Journal.*

Tijssen, Robert J. W., Ludo Waltman, and Nees Jan van Eck. 2011. "Collaborations Span 1,553 Kilometres." *Nature* 473 (7346): 154–154.

Torre, André, and Delphine Gallaud. 2004. "Geographical Proximity and Circulation of Knowledge through Inter-Firm Cooperation." In *ERSA Conference Papers.* Econ Papers.

Townsend, A. M., S. M. DeMarie, and A. R. Hendrickson. 1998. "Virtual Teams: Technology and the Workplace of the Future." *Academy of Management Perspectives* 12 (3): 17–29.

Watts, Duncan J, and Steven H Strogatz. 1998. "Collective Dynamics of 'Small-World' Networks." *Nature* 393 (June): 440–42.

Wineman, J. D., F. W. Kabo, and G. F. Davis. 2008. "Spatial and Social Networks in Organizational Innovation." *Environment and Behavior* 41 (3): 427–42.

Wineman, Jean, Yongha Hwang, Felichism Kabo, Jason Owen-Smith, and Gerald F Davis. 2014. "Spatial Layout, Social Structure, and Innovation in Organizations." *Environment and Planning B: Planning and Design* 41 (6): 1100–1112.

Wuchty, S., B. F. Jones, and B. Uzzi. 2007. "The Increasing Dominance of Teams in Production of Knowledge." *Science* 316 (5827): 1036–39.