Internal Network Structure and the Speed of Generative Appropriability

Nicholas Argyres
Olin School of Business
Washington University in St. Louis

Luis Rios
Krannert School of Management
Purdue University

Brian S. Silverman
Rotman School of Management
University of Toronto

December 2022

Abstract

Strategy research has long been concerned with how firms protect and build on the technological knowledge they create, and has focused on legal enforcement, complementary assets, and location decisions. Less attention has been paid to how internal firm structures support the appropriation of future cumulative innovations: what has been termed "generative appropriability." Drawing on innovation and social network research, we propose that more connected intrafirm inventor networks facilitate generative appropriability by accelerating within-firm generation of follow-on innovations, thus pre-empting rivals. Using patent data on 1,391 large corporations over 26 years, we find that more connected internal inventor networks are associated with more and more valuable cumulative patents generated by focal firms relative to other firms. This effect is strongest in the critical first few years after an initial invention.

We thank seminar audiences at Cornell University, INSEAD, University of California – Irvine, University of Southern California, University of Manchester, University of Michigan, University of Wisconsin, and Washington University in St. Louis. Rachel Justine Smith provided excellent research assistance.

INTRODUCTION

Proprietary technological knowledge often forms the basis of a firm's competitive advantage. Indeed, according to the knowledge-based view of the firm, the primary rationale for the firm as an institution is to generate and integrate such knowledge (Kogut & Zander 1993; Spender 1996; Nahapiet & Ghoshal 1998). However, given the non-rivalrous nature of knowledge (Arrow 1962) and the cumulative nature of technological development (Dosi 1984), firms must not only find ways to build on the knowledge they generate but also outcompete others in the effort to do so (Teece 1986). That is, they must seek to maximize what Ahuja, Lampert and Novelli (2013) conceptualized as the "generative appropriability" of their innovations: "a firm's effectiveness in capturing the greatest share of future inventions spawned by its existing inventions" (p. 248).

Appropriability (generative and otherwise) has been amply studied in the context of external institutional and market factors. It often relies on formal legal mechanisms such as patents, trade secrets, and employee non-compete agreements (Cohen, Nelson & Walsh 2000; Marx, Strumsky & Fleming 2009; Somaya 2012). Locating in sparse geographic or technological spaces may also reduce the rate of knowledge spillover to rivals (Alcacer & Chung 2007; Bloom, Schankerman & van Reenan 2013). An earlier literature studied the role of downstream advantages such as complementary assets and early product market entry in appropriation of returns to innovation (e.g., Teece 1986; Lieberman & Montgomery 1988; Mitchell 1991; Tripsas 1997; Rothermael 2001). In contrast, the role of internal organization in affecting appropriability has been largely overlooked. A notable exception is Liebeskind (1996), who proposed that job design can be used to reduce knowledge leakage by dividing tasks so that few employees possess sufficient knowledge to leak, but also warned that such moves might also impinge on organizational integration mechanisms that are needed to build cumulatively on the firm's knowledge (Grant 1996; Kogut & Zander 1996).

Another exception is Zhao (2006), who found evidence that multinational firms protect the technologies they develop in countries with weak legal intellectual property by using them more internally, and by linking them more strongly to other technologies of the firm.

We extend this research stream by investigating how internal structure might help a firm protect and build on its own knowledge by accelerating the rate of its generative appropriability. Our core argument consists of two parts: First, drawing on prior work, we highlight how in a competitive setting it is not enough for a firm to develop technologies that build upon its prior innovations—it must do so *quickly* in order to foreclose follow-on innovation by its rivals; and second, we propose that firms with more connected intra-firm networks can build on their own technologies faster, thereby improving generative appropriability. In a more connected internal network, it is increasingly likely that unrelated inventors (those not involved in the original invention itself) may be directly or indirectly exposed to the invention's underlying knowledge. Inventors in connected networks may have an advantage in the quest to rapidly pursue follow-on research along multiple trajectories—a quest that is embarked upon by the focal firm and its rivals. We thus build on the literature relating network structure to innovation (e.g., Reagans & McEvily 2003; Sorenson, Rivkin & Fleming 2006; Grigoriou & Rothermael 2017; Moreira, Markus & Laursen 2018). Although this research has shown that firms vary significantly in the structure of their internal inventor networks – ranging from highly connected to highly fragmented – and that these differences matter for innovation diffusion, absorption and impact, to our knowledge no work to date has explored the role of structure on the timing of generative appropriability.

To empirically test these novel propositions, we measure the internal inventor network connectedness and patenting output of 1,391 large American corporations from 1986 to 2019. We find evidence consistent with our argument that greater network connectedness among a firm's inventors enables a focal firm to pre-empt the technological space surrounding its innovations, amassing cumulative patents faster for a given invention relative to its rivals. We also present evidence that such accumulation of patented knowledge is associated with greater appropriation of the returns to innovation, consistent with Ahuja et al.'s (2013) proposition that patent self-citation ratios are a reasonable proxy for generative appropriability.

LITERATURE AND HYPOTHESES

Knowledge, imitation, and the speed of generative appropriability

Strategy scholars have long emphasized that a firm's competitive advantage often derives from the development of unique and valuable knowledge – particularly of a technological nature– that is difficult to imitate (Spender 1996; Kogut & Zander 1996). In this view, firms are seen as institutions that serve to coordinate and integrate diverse and specialized knowledge stocks, which are often held by individual employees (Grant 1996; Barney 1986). The literature on dynamic capabilities similarly highlights the firm's ability to continuously upgrade its knowledge base in response to environmental changes (Teece, Pisano & Shuen 1997).

Given the centrality of technological knowledge to competitive advantage, protection of that knowledge is paramount. This is because rivals constantly seek ways to build on the knowledge created by other firms. Ahuja et al.'s (2013) construct of generative appropriability highlights the importance of "a firm's effectiveness in capturing the greatest share of future inventions spawned by its existing inventions". Thus, it is not enough for a firm to build upon its knowledge, it must do so more effectively than its rivals to prevent expropriation. Consistent with this emphasis, much work has focused on strategic, legal and institutional mechanisms of protection, such as location choices (e.g., Alcacer & Chung 2007); patent enforcement efforts (Agarwal, Ganco & Ziedonis 2009); confidentiality and employee non-compete agreements (Marx et al. 2009); and job designs (Liebeskind 1996; Rajan & Zingales 2001.

Recent work, however, has also called attention to subtler internal mechanisms available to firms in the pursuit of knowledge protection, such as the ways in which knowledge is manifested in prototypes, formulas, routines, blueprints, and the like (Sharapov & Macaulay 2022) . One vivid example of this kind of protection strategy is Michelin Tire Company's secret firm-specific recipes for curing tires – a process that required very specific temperature and pressure specifications. In the latter half of the 20th century, Michelin's machines had thermostats that were calibrated to a unique Michelin scale whose conversion to Fahrenheit or Celsius degrees was known only to a handful of top executives. Thus, a Michelin employee who was hired by a rival tire company could only say "we cure the tires for 30 minutes at 50 degrees Michelin," which was not very informative.[1]

More generally, organizational structure can shape a firm's generative appropriability by influencing the speed of follow-on invention. As Kogut & Zander (1992: 393) note, knowledge-based competition can be characterized as "a race between an innovator and the ability of the imitating firm [to] reverse engineer…the substantive technology. The growth of the firm is determined by a combination of the speed of technology transfer and of the imitative efforts of rivals." Thus, to the extent that a firm is better able to rapidly build upon its own innovations, it will more successfully pre-empt rivals from doing so. In the next section, we propose how structure can facilitate knowledge flow and in turn increase the speed of generative appropriability.

Inventor network structures and generative appropriability

We still know little about how internal inventor network structures may affect generative appropriability through their impacts on knowledge diffusion. This is an important gap because managers can affect these structures through various organizational choices (Argyres, Rios &

[1] Personal communication between one of the authors and a former Michelin executive.

Silverman 2020). While several studies show how within-firm inventor networks affect the diffusion of knowledge within the firm,[2] they are silent regarding how internal knowledge flows interact with flows to rivals. At the individual level, Singh (2005) finds evidence that inventors' knowledge is more likely to flow to someone with whom they have previously co-invented than to a non-collaborator. Sorenson, et al. (2006) similarly find that an inventor's knowledge is more likely to diffuse to socially proximate inventors (as measured by co-invention), and that this diffusion rate decays slowly with degrees of separation between two inventors in a co-invention network. Singh & Agrawal (2011) find that the knowledge of a newly-hired inventor diffuses more readily to incumbent inventors who collaborate with the inventor than to non-collaborating inventors (see also Tzabbar, Silverman & Aharonson 2015). At the organizational level, the literature also shows that knowledge diffuses through inventor networks, suggesting that more connected networks may spur the development of follow-on innovations necessary for exploiting earlier innovations. For example, a few studies have analyzed the impact of internal inventor networks on the extent and kinds of innovations generated and absorbed (e.g., Guler & Nerkar 2012; Argyres et al. 2020). In particular, Carnabuci and Operti (2013) find that firms with more integrated inventor networks and a more diverse knowledge base tend to mitigate the trade-off between recombining new knowledge vs. recombining prior knowledge, while Moreira et al. (2018) show how such structures can speed the absorption of external knowledge.

. Consistent with this, interviews of pharmaceutical R&D managers suggest that the absence of linkages among a firm's inventors can prevent a firm from building upon its own knowledge:

[2] The antecedents of firms' internal inventor network structures have not been studied *per se*, although factors such as the firm's geographical profile (e.g.,Singh & Marx 2013), the nature of its parent-subsidiary relationships (e.g., Birkinshaw & Hood 1998), and its knowledge transfer strategies (e.g., Winter & Szulanski 2001) likely influence them. Moreover, because a firm's internal network structure is likely to change faster than its geographic profile or parent-subsidiary relationships, it is unlikely that the relationship between internal network structure and generative appropriability is driven by such factors. Our analysis is therefore agnostic about them.

> "*...an idea in one area may be able to be translated into another therapeutic area. Quite often an indication may be unsuccessful in one therapeutic domain but have legs in another, however with the wrong structure (they) may not be able to take advantage of this.*" (Balachandron & Eklund 2019)

A firm's ability to achieve generative appropriability, however, depends not only its own internal efforts, but also the speed of those efforts relative to rivals' expropriation efforts. The practitioner literature has documented many examples of firms that failed to quickly capitalize on their own technologies, only to watch rivals leapfrog them and find new uses for inventions that appeared to be exhausted by the original inventing firm. A notable example is how Pfizer found a way to redeploy Boceprevir to their own blockbuster COVID treatment drug Paxlovid. Boceprevir had been developed by Merck decades ago to fight hepatitis C, but Merck was unable to quickly identify its potential in fighting COVID. It is estimated that Pfizer will realize $17 billion from this drug in 2022, while Merck watches from the sidelines (Lauerman 2022). Such cases highlight the importance of understanding how firms can fail to exploit technologies in which they should have had a lead over rivals. However, with the exception of Zhao's (2006) study of patent protection mentioned above, we are unaware of any empirical literature relating internal organizational structure to generative appropriability. Thus, while we build on the above-mentioned studies which focus exclusively on knowledge development within firms, our goal in this paper is to study firms' appropriation of knowledge *in relation to others' attempts to do so*. This is of course the key determinant of generative appropriability. Our hypothesis development proceeds in two steps. First, we propose a baseline relationship between structure and the level of generative appropriability, then seek to unpack the role of speed in enhancing generative appropriability.

Hypotheses

What kinds of information pass through a firm's internal inventor networks to facilitate the process of building on prior innovations? We suggest that there are at least three types. First, information about the existence of new innovations on which the firm can build may flow through such networks, making inventors in disparate locations or subunits of the firm aware of new opportunities to exploit such innovations before rivals do. The role of social networks in transmitting information about such innovation opportunities has long been emphasized by network scholars (e.g., Reagans & Zuckerman 2001; Obstfeld 2005). Second, tacit or partly tacit knowledge may flow through internal inventor networks. This kind of uncodified knowledge is excluded from patent applications or invention descriptions, and hence unavailable to rivals, and interpersonal communication may facilitate building on the original innovation (e.g., Polanyi 1962; Teece 1982; Winter 1987). Finally, a large literature emphasizes the importance of learning from failures in general (e.g., Nelson & Winter 1982; Sitkin 1992; Cannon & Edmonson 2005), and from failed R&D projects in particular (e.g., Eggers 2012; Khanna, Guler & Nerkar 2016). Thus, information about prior failed innovative efforts related to an innovation may be transmitted through internal inventor networks. Such information is not publicly disclosed, yet may be extremely valuable in guiding efforts to build upon an initial innovation by enabling researchers to avoid dead-ends in the research and development process. The foregoing highlights the importance of pursuing a better understanding of what features of intrafirm networks may enable faster diffusion of information within the firm, relative to its diffusion to its rivals.

Given the demonstrated role of network ties as conduits of information, firms with more connected internal inventor networks should enjoy more robust flows of these three types of information across a wide range of inventors. They will therefore more efficiently build on their existing innovations, pre-empting rivals from doing so. Thus we predict:

Hypothesis 1: The more connected a firm's internal inventor network, the greater level of generative appropriability it will achieve.

Despite the availability of numerous mechanisms for improving generative appropriability, all but the most complex and/or tacit technological knowledge eventually diffuses across organizational, technological, and geographic boundaries (Jaffe, Henderson & Trajtenberg 1993; Jaffe & Trajtenberg 1999; Sorenson et al. 2006). Studies of innovation have repeatedly demonstrated that rival firms eventually manage to reverse engineer or "invent around" valuable patents (Mansfield 1961; Mansfield, Schwartz & Wagner 1981; Cohen, Nelson & Walsh 2000). Therefore, inventing firms have a limited time window within which to build on their own valuable knowledge before rivals begin to do so (Winter 1987; Ceccagnoli 2009), effectively racing to recombine their knowledge internally before others can successfully build upon their discoveries (Kogut & Zander 1992).

Because time is of the essence, and because prior literature demonstrates that local knowledge diffuses over time, we contend that a firm's efforts to build on its own knowledge will be most effective shortly after the disclosure of an initial innovation, and that this advantage should decline over time. Ideally from the focal firm's perspective, by the time the knowledge underlying the initial innovation fully diffuses to rivals, it will have introduced follow-on innovations and begun efforts to build upon them, thereby pre-empting expropriation by those rivals. Therefore, we hypothesize that:

> *Hypothesis 2: The positive relationship between the connectedness of a firm's internal inventor network and generative appropriability will be strongest immediately after the initial invention, and will weaken as the initial invention ages.*

DATA AND MEASURES

Our data are derived from three sources: the European Patent Office's (EPO's) PATSTAT database, Bureau van Djik's (BvD's) ORBIS database, and COMPUSTAT. We collected patent data for the time period 1986 through 2019 from the PATSTAT database, which features perhaps the most comprehensive global coverage of patenting in existence. We linked patents to their original owners using ownership structure data from ORBIS. The extensive linking procedure was based on prior matching completed in a collaboration between the EPO and BvD, following Rios

(2021). This procedure served to eliminate acquired and divested patents from the analysis, since such patents do not measure the firms' innovation outcomes, nor can they be used to measure internal inventor networks. Inclusion of such patents would introduce bias, for example, if a firm sold a patent shortly after grant. If divestment of a patent indicates that a firm has little interest in further pursing that technology, then a patent sale will be associated with little follow-on innovation by the firm. In our estimations, we would misinterpret a lack of follow-on development as failure to appropriate value, when in fact the value was extracted via a sale. In addition, this procedure allowed us to link to corporate parents any patents assigned to their wholly owned subsidiaries. This is important because patents assigned to subsidiaries might be mistaken for external patents, even though they were generated within the firm (Arora, Belenzon & Rios 2014).

Consistent with prior work, we limit our sample to firms with more than 250 U.S. patents over the sample period to ensure that we focus on firms for whom patenting is an important economic activity. We also limit the sample to firms that appear in COMPUSTAT, so that we can control for relevant firm characteristics. The resulting panel matches all patents, applications, and reassignments between 1986 and 2013 for the 1,391 firms in our sample and their wholly owned subsidiaries, and allows us to record citations, renewals, foreign applications, divisionals, and continuations to these through 2019. We collected firm-level data from COMPUSTAT for these firms for the same time period.

We adhere to conventional practice in our use of patent data, with the exception of adopting two improvements to traditional variable construction. First, rather than treating each patent as a stand-alone entity, we follow recent work that aggregates patents to the domestic priority family level (e.g., Kuhn, Younge & Marco 2020; Rios 2021). Patent families are groups of patents and ungranted patent applications that all trace back to the same initial priority application and thus capture the full impact of a technological advance more clearly than a simple tabulation of patents

(Martinez 2011). Recent research demonstrates that firms utilize the patent system strategically, often resulting in many granted patents covering different aspects of the same underlying innovation (Kuhn, Younge & Marco 2020). Figure 1 provides an illustrative example from the patents earned by Square, Inc. The seminal technology behind Square's payment system spawned a large domestic family of continuations that claimed the earlier priority date and covered the exact technical content (i.e. drawings and descriptions) but refined and extended the claims or application of the technology. Thus, domestic families can be thought of as a bundle of legal rights that encapsulate one invention, rather than multiple independent inventions. Aggregating forward citations at the level of the domestic family is therefore a more accurate way of capturing the impact that a firm's invention has on the internal and external innovation landscape.

There are two categories of patent families: jurisdictional and domestic. For jurisdictional families, the same invention may gain patent protection in different countries, often under the Paris Convention, whereby member countries recognize priority protection of patents filed in other jurisdictions. However, for domestic families, which only apply to US patents, a mechanism exists whereby an initial application or "priority" can both grant early protection and also allow the inventor to further refine the technology later. This way, an invention may be amended by adding new claims (a "continuation") or can be split into multiple patents, each with narrower claims (a "divisional"), while still maintaining the protection afforded by the earlier priority filing date. This is important in our setting because prior work has used citations to individual patents without considering the fact that many patents essentially cover the same technology in this way. Consistent with our focus on U.S. patents, we aggregate domestic patent families and exclude jurisdictional families.

Second, we include citations to patent applications as well as to granted patents. Patent research has traditionally included only citations to already-granted patents, largely because these data were easy to collect. However, because patent applications are published 18 months after submission, a large number of citations are to patent applications that have not yet been granted. Kuhn et al. (2020) demonstrate that the inclusion of such citations substantially challenges much "received wisdom" in the literature. Given our focus on the value of rapid follow-on innovation and knowledge protection, such citations are likely to be particularly important. Figure 2 illustrates the potential importance of citations to applications. The bottom panel shows the patent application for an invention related to Google Glass, which was published on March 9, 2006. The USPTO ultimately granted the patent on April 13, 2010 (top panel). The ungranted patent application received 97 citations (all prior to April 13, 2010), while the granted patent has received 42 citations through June 2022. Thus, omitting citations to patent applications can change the interpretation of an innovation's impact, and particularly important for this study, fail to identify follow-on innovations that are developed quickly after an initial innovation.

Table 1 presents definitions of all variables. We elaborate on these below.

[INSERT FIGURES 1 AND 2, AND TABLE 1, ABOUT HERE]

Dependent variable

Self-citation ratio. Our primary measure of generative appropriability is a patent-family's forward self-citation ratio. As Ahuja et al. (2013: pp. 266) note, "A simple archival measure of GA [generative appropriability] is provided by the ratio of self-citations to a firm's patents to the total citations received by the firm's patents." Their logic follows much work that views citations as a reasonable albeit noisy proxy for cumulative innovation (Furman & Stern 2011; Galasso & Schankerman 2015). Thus, self-citations reflect a firm's appropriation of the value of its inventions through follow-on inventions, while citations by others reflect knowledge leakage (expropriation). The ratio of self citations relative to others' citations therefore captures generative appropriation.

This measure of generative appropriability goes well-beyond the simple self-citation variables available on various extant datasets, An empirical innovation in our study is that we are able to calculate the self-citation ratio at the patent family level.[3] We are thus able to directly observe the timing of how the knowledge associated with each invention is incorporated into the focal firm's subsequent patents, as well as those of its rivals. A citation included in a patent filed by firm B to a patent filed by firm A is generally viewed as evidence that firm B has built upon firm A's innovation in some manner (Jaffe, Trajtenberg & Henderson 1993; Jaffe & Trajtenberg 2002). Patent self-citations are similarly thought to reflect a firm's effort to build on its own prior inventions, and therefore our self-citation ratio is conceived as a scale-free reflection of how much a firm builds upon its own innovations relative to how much other firms do. Consistent with this assumption, Wagner and Wakeman (2016) found evidence that self-citations are associated with the commercial success of new pharmaceuticals. Also consistent is Moser, Ohmstedt & Rhode's (2018) finding that versions of hybrid corn created to generate valuable follow-on versions received nearly 50% more self-citations than hybrids that were not generative in this way. [4] We formally define $SelfRatio_{jk}$ as the count of all of firm *k's* self-citations to patent family *j* (i.e., forward citations that appear in firm *k*'s subsequent patent applications) divided by the count of all forward citations by any firm to that patent family. We exclude citations that were added by patent examiners as they

[3] It is important to note that prior work has taken a pooled approach to self-citations (always at the firm level). There are alternate ways to calculate self-citation ratios, depending on the question at hand. For example, Hall, Jaffe & Trajtenberg (2001) measured it as follows: "for each patent that has an assignee code we count the number of citations that it made to (previous) patents that have the same assignee code, and we divide the count by the total number of citations that it made." However, those authors also defined self-citations as "those coming from down-the-line patents owned by the same firm [vs] citations coming from external firms" (Hall, Jaffe & Trajtenberg, 2001). The authors thus used the term "self-citation" to mean either the ratio of backward citations to the firm's own patents to its total citations, or the ratio of backward plus forward citations to the firms' own patents to its total citations. Subsequent work has also measured self-citation ratios in these alternate ways. As stated earlier, our measure is unique in capturing the ratio at the patent-family level, rather than aggregating at the firm-year level.

[4] Early research on patent citations suggested that citation counts were generally valuable to the firm (Scherer 1956; Grilliches 1990), but later work has found that various moderators can reduce their association with value (Schankerman & Pakes 1986; Hall & Harhoff 2012; Abrams, Akcigit & Grennan 2018). Hall et al. (2005), for example, found that citations were associated with firm value (measured by Tobin's Q), but for small firms only. Firms' strategic use of the patent system may be one reason that citations are not always valuable (Kuhn et al. 2020).

may be added for reasons other than knowledge flow (Alcacer & Gittelman 2006), and citations among direct co-authors or from inventors to themselves, since these reflect mechanical connections rather than network channels of knowledge flow. By construction, since our measure is at the patent-family level, we do not include citations from any patent to any "member" of its family (e.g., earlier priorities or subsequent continuations or divisionals), since these are not cumulative innovations, but rather legal extensions of the same invention (Kuhn et al, 2020). For our self-citation ratio to be a credible measure of generative appropriability and appropriability more generally, it should be associated with a firm's ability to extract value from its patented innovations. Thus, a set of post-hoc empirical tests in the next section provide evidence that our self-citation ratio is positively associated with four different measures of patent value.

Independent variables

Intrafirm inventor networks. Our measures of the structure of a firm's intrafirm inventor network consider each inventor in the firm to be a node, and are calculated based on non-directional ties that are recorded each time an inventor collaborates on a patent application with another inventor within the firm.[5] These ties are observed separately for each firm-year, and are based on patent applications filed in that year.

Inventor networks may be more or less connected for numerous reasons. Extant literature would suggest that clusters may exist because of geographic distance (Singh & Marx 2013) or autonomous subsidiaries (Arora et al, 2014). For the purposes of this study we are agnostic as to the determinants of structure—put simply, whether a group of inventors is isolated from the core network of inventors due to geography or hierarchy should not change the fact that this isolation is

[5] Note that the number of inventors per patent is immaterial with regard to our dependent variable (self-citation ratio) because citations are measured at the patent level and because we exclude forward citations made by a focal patent's own authors.

likely, on average, to reduce the potential for knowledge transfer. The literature on social networks has tended to take such networks as exogenous, and has only recently begun to investigate the determinants of network structures (e.g., Ahuja, Soda & Zaheer 2012).

While it may seem straightforward to think of a network as more or less connected, formal measures of network connectedness have many dimensions and can be conceptualized in several different ways. The utility of each concept depends on the nature of the data and the phenomenon being investigated. At a whole-network level, a network is maximally connected if there is a path between every pair of nodes (Wasserman & Faust 1994), and becomes less connected with each additional isolated "component" in the network, where a component refers to a node or group of nodes disconnected to any other group (Carnabucci & Operti 2013). Here, the largest component is termed the "giant" component (Bollobás 2001). Conventionally, a majority of studies that look at inventor networks analyze the properties of this central component, because standard ego-network measures necessitate that there be at least one edge connecting all nodes in a network (e.g., Nerkar 2005). This is informative when the giant component encompasses most of the nodes in a network. However, as the number of nodes located in isolated components grows, the network becomes less connected (more fragmented), and network measures that focus solely on the structural features of the giant component lose appeal because they ignore the topography of the peripheral subnetworks. In our sample roughly 60% of patents are generated by inventors in peripheral components, so focusing on the properties of the giant component alone would not capture the richness of our networks. We therefore employ two increasingly popular measures of whole-network connectedness: Normalized entropy ($Entropy_{kt}$) and relative giant component size ($Giant_{kt}$).

Prior studies have used the size of the giant component relative to the size of the network – that is, the number of nodes connected in the largest component relative the maximum number that could be connected – as a measure of connectedness in the network (e.g., Argyres, et al, 2020;

Kogut, Urso & Walker, 2007). Such information is useful to know the extent to the proportion of nodes that are part of the main network. For example, a very large relative giant component in an inventor network would tell us that most inventors are connected to each other, at least indirectly. An alternative measure, termed "normalized entropy," complements measures of the giant component by incorporating information about the peripheral subnetworks (e.g., Amburgey 2008; Amburgey, Al-Laham, Tzabbar & Aronson 2008). Normalized entropy captures the degree of equality in the size of a network's components (both giant and peripheral), and is conceptually similar to other concentration and diversity measures such as the Herfindahl–Hirschman Index (HHI) and the Shannon Diversity measure. Networks with high normalized entropy have components that are of relatively equal size, whereas lower entropy values indicate variation in size of components. Thus, a high entropy value would indicate that the components are all relatively small and thus would have a very low "critical mass" of inventor connectedness among all the sub-networks for a firm; in contrast, lower entropy values would indicate the presence of one or more larger components with higher critical mass of connectedness. When thinking about the knowledge recombination potential for a fixed number of inventors, the normalized entropy measure captures the impact of structure on the number of possible connections in a mechanical way following the logic of Metcalfe's law of network connections; namely, an inventor who moves from a smaller component to a larger one increases the number of possible connections with other inventors. Figure 3 illustrates this phenomenon by considering an idealized firm consisting of sub-networks.

[INSERT FIGURE 3 HERE]

Normalized entropy is our preferred measure of connectedness given the additional information that it provides, but for completeness we employ both measures below.

Entropy is thus measured as:

$$H = -\sum_{c=1}^{C}[(Nc/N) \ * \ log(Nc/N)] \ , \tag{1}$$

where C is the number of components and N_c is the number of nodes in component c. Since the maximum value that H can take is logC, we can normalize its value between 0 and 1. *Normalized entropy* equals 1 when all components have the exact same size, and is calculated for firm k, year t:

$$Entropy_{kt} = H_{norm} = \frac{H}{logC} \tag{2}$$

Note that this measure of normalized entropy measures the *inverse* of network connectedness. In other words, higher values of Entropy are associated with lower levels of network connectedness. Thus, for ease of interpretation, we reverse-code Entropy so that higher levels of the variable are predicted to be positively associated with self-citation ratio. Put simply, in our econometric specifications we multiply our measure of Entropy by -1 to create *normalized negative entropy,* which proxies for connectedness.

Our alternative measure of network structure, *relative giant component size*, is formally measured as the proportion of a firm's inventors who are connected in the largest cluster within the firm's co-invention network. Specifically:

$$Giant_{kt} = \frac{NumInvLargestComponent_{kt}}{NumInvFirm_{kt}} \tag{3}$$

where $NumInvLargestComponent_{kt}$ represents the number of inventors in the largest component of firm k's network in year t, and $NumInvFirm_{kt}$ represents the total number of inventors at firm k in year t. The numerator and denominator are based on all patent applications that firm k submits in year t. As explained above, relative giant component size is a less appropriate measure of network structure for our particular empirical context because it does not capture the structure of the periphery components. Nonetheless, we include it in order to contextualize our paper within the aforementioned series of studies that have used this measure.[6]

It is important to emphasize that our independent variables $Entropy_{kt}$ and $Giant_{kt}$ capture potential channels for knowledge flow between inventors that have at least one, and usually many, degrees of separation between them. This reflects the large sizes of the publicly traded firms in our sample, as they employ often thousands of inventors. Thus, even in what would be considered a highly connected structure (low normalized entropy and large giant component), knowledge that may be useful for building on any given invention may lie anywhere along the extensive paths between inventors in the firm who are indirectly connected to each other. These paths involve co-authorship links not only between any first-degree pair downstream in the network, but between any such pair that can be traced back to original inventors through intermediate connections. Indirect co-authorship connections in academia provide an analogy. A focal scholar may have never worked with, or even met, a certain other scholar, but the two may be indirectly connected through a third co-author. Similarly, many of the within-firm citations to a focal patent family that we measure are made by inventors who are have never collaborated with the focal patent family's authors. Given this, and given that we exclude forward citations by the focal patent's inventors, the relationships we estimate between network structure and the self-citation ratios cannot merely be the result of a mechanical correlation between our dependent and independent variables.

Patent value

[6] As an alternative, one might consider aggregating the ego-network characteristics of a firm's inventors to calculate average measures. However, these fail to incorporate information on the topology of the whole network. For example, a firm with equal-sized components and another firm with components of widely varying size might exhibit the same average density of collaboration, while having extremely different contexts for collaboration. As well, a "small worlds" measure of network connectedness (e.g., Baum, Shipilov & Rowley 2003; Schilling & Phelps 2007) would be based on navigability within a network. Because it is impossible to navigate among the (many) disconnected components that we observe in real-firm networks, such a measure is not useful for the type of analysis we pursue.

As referenced above, Ahuja et al. (2013) endorse self-citation ratio as a valid proxy for generative appropriability. However, there is the possibility that self-citations may not reflect value; in other words, appropriation of a greater share of cumulative patents may not translate into appropriation of more returns to the focal innovation. In post-hoc analyses, we estimate the effect of the self-citation ratio on various measures of the private value of a firm's patent family. This purpose of this exercise is to assess the extent to which *SelfRatio* does in fact proxy for successful appropriation of returns. We construct four measures of value. Our first proxy is based on patent renewals. The U.S. Patent and Trademark Office (USPTO) charges a renewal fee at certain points during a patent's life.[7] The patent assignee has the option to pay this fee in exchange for continued patent protection, or to forgo paying this fee and allow patent protection to lapse before the patent's maximum possible 20-year life expires. A long line of research has used patent renewals to infer the value of a patent because firms are presumably only willing to pay renewal fees if the patent's value exceeds the value of the fee (e.g., Pakes 1986; Serrano 2018). Therefore, we calculate $YrsRenewed_{jk}$, measured as the average number of years that patents in patent family j were renewed by firm k.

A second proxy for patent value is the number of countries ("jurisdictions") in which a firm applies for patent on a given invention. While it is well known that most patents are not commercially valuable, a firm with private knowledge about the potential for a technology will seek to protect it in more international jurisdictions in proportion to its assessment of quality (Wagner & Wakeman 2016). International filings are not trivial in terms of cost, given the need to translate and conform to various different patenting regimes in addition to the actual filing fees. Each international patent can easily cost over $20,000. $Jurisdictions_{jk}$ is a count of the number of countries in which the firm filed applications for a given family. A third alternative measure of

[7] The fee schedule for large firms such as those in our sample is as follows: $1,600 due 3.5 years after a patent is granted, $3,600 due 7.5 years after granting, and $7,400 due 11.5 years after granting.

patent value is $Triad_{jk}$, a dummy for whether the family included filings in the three major patent jurisdictions: US, EPO, and Japan.

Our final proxy of patent value is $AvgPatentFamilySize_{kt}$. As noted above, priority patent families consist of groups of patents within the same jurisdiction (in this case the USA) that all relate to the same underlying innovation. Often, patent families include a core, or "productive" patent, along with surrounding "defensive" patents (Abrams et al. 2018). Average patent family size reflects the value of productive patents, because firms are presumably willing to invest in a greater number of defensive patents in order to protect more valuable productive patents. Family size and number of jurisdictions are correlated but not identical, because a firm may file multiple applications within multiple jurisdictions, so these are two dimensions indicating the value and exploitation of a seminal idea.

Control Variables

Patent-level measures. We control for several characteristics of the underlying innovation that might affect subsequent citation patterns. This is important because prior work has found associations between inventor network connectedness and the quality of inventions generated (e.g. Argyres, et al. 2020). Thus, it may be that differences in citation patterns could be spuriously driven by different types of patents being generated by more connected firms. To control for the breadth of search underlying an innovation, we construct the variable $Originality_{jk}$ for each patent family j (Hall et al. 2001), calculated as one minus the Herfindahl index of the primary U.S. patent classes of the patents cited by each patent in patent family j.

Similarly, we measure the breadth of innovation impact by constructing the variable $Generality_{jk}$, calculated as one minus the Herfindhal index of the primary U.S. patent classes of the patents that cite a patent (Hall et al. 2005). $NonPatentReferences_{jk}$, measured as a count of the references to non-patent literature (i.e., scientific publications) made by each patent, controls for the "basicness" of the research underlying a patent. This variable reflects the extent to which a firm is on the technological frontier, which also may affect the propensity of its patents to be cited. Finally, a firm's ability to protect its knowledge may depend on knowledge complexity, because more complex knowledge is more difficult to imitate (Rivkin 2001). We therefore control for $PatentComplexity_{jk}$, the complexity of the knowledge represented by a patent family, by incorporating the complexity measure developed by Fleming & Sorenson (2004) and Sorenson et al. (2006). This proxy for complexity incorporates historical information on the frequency of recombination of each patent subclass, then creates a patent-level measure of interdependence by averaging across all subclasses represented within a patent.[8] These patent family-level measures are then normalized within a family's technology class and within their application year in order to mitigate arbitrary variance due to time trends or technology characteristics.

Firm-level measures. Firm size may influence the nature of innovation undertaken (Cohen & Klepper 1996). We therefore include $LnSales_{kt}$, $LnAssets_{kt}$, and $LnEmployees_{kt}$, measured as the natural log of sales, assets, and employees, respectively, to control for such size effects. A firm's R&D expenditure is also likely to affect its innovative efforts. We therefore include $LnR\&D_{kt}$, measured as the natural log of R&D expenditure. Other aspects of a firm's research effort may affect its innovative efforts. We include $LnPatentCount_{kt}$, which captures the natural log of the number of ultimately successful patent families that firm *k* applied for in year *t*. In estimations of network properties, we also control for the size of the network via $NumComponents_{kt}$, measured as the number of separate components in the co-invention network. All firm-level measures are lagged by one year.

[8] We construct the complexity measure using CPC codes, whereas prior work used the older IPC coding scheme. Details on the construction of the measure on Table 1.

Fixed effects. In our patent-level regressions we include firm, technology class, and year fixed effects, (firms rarely change primary NAICS, so we do not include industry fixed effects). Technology class and year fixed effect mitigate the concern that industry life-cycles or secular trends would drive observed variation.

[INSERT TABLES 2 AND 3 ABOUT HERE]

Tables 2 and 3 report summary statistics and conditional correlations for our data. As Table 3 indicates, *Entropy* and *Giant* are highly negatively correlated (r = -0.97). This is not particularly surprising given that a larger *Giant* is likely to increase the variation in the size of a network's components. Given this correlation, we do not include both variables in the same model below. Not surprisingly, the correlation among assets, sales, and employees is also high, since larger firms tend to have more sales and employees. However, we include all three of these variables because firms in some technology-based industries often have lower employee-to-asset ratios. The results are not sensitive to alternative specifications involving these three measures.

METHODS AND RESULTS

Organizational Network Structure and Self-Citations

We first explore the relationship between self-citation and inventor network structure. Hypothesis 1 predicts a positive relationship between the connectedness of a firm's inventor network and the extent of generative appropriability for its innovations. We examine this using our proxies for network connectedness – *Negative Entropy* and *Giant* – and for generative appropriability – *SelfRatio*. Following convention in the literature (e.g., Hall et al. 2005), and in order to make our results easier to interpret and relate to prior work, our first specification employs ordinary least-squares regressions of the form:

$$SelfRatio_{jk} = \alpha + \beta_1 Networkmeasure_{kt-1} + \gamma X_{kt-1} + \theta Z_{jk} + \delta Firm_k + \omega Year_t + \mu Tech_{jk} + \epsilon_{jk} \qquad (4)$$

where the independent variable *Networkmeasure* is *Negative Entropy* and *Giant*; *X* is a vector of time-varying firm-specific covariates for firm *k* at year *t*-1; *Z* is a vector of patent-specific covariates averaged across all the applications in family *j*; and *Firm, Year* and *Tech* are fixed effects (the latter two being the year of application and the technology class for the initial priority patent application for family *j*). The coefficient of interest is β_1, which reflects the relationship between network structure and outcome.

INSERT TABLE 4 ABOUT HERE

Table 4 presents our core results. Models 1-4 regress *SelfRatio* on *Negative Entropy*, (recall that we convert our measure of entropy by multiplying by -1 to ease interpretation) introducing a progressively larger set of control variables. In Model 1, the coefficient on *Negative Entropy* is 0.125 ($p < 0.000$); a one-standard-deviation increase in the relative size of *Negative Entropy* is associated with an increase in *SelfRatio* of 0.020. Given the mean of the dependent variable (0.163), such an increase in *Negative Entropy* suggests a commensurate increase of more than 12% in *SelfRatio*. The coefficient on *Negative Entropy* remains nearly identical as controls are added. Models 5-8 regress *SelfRatio* on *Giant*, again with increasingly large sets of control variables. The coefficient on *Giant* is stable across all models (it is approximately half the magnitude of the coefficient on *Negative Entropy*) and positive as expected ($p < 0.003$). A one-standard-deviation increase in *Giant* is associated with an increase in *SelfRatio* of 0.012, or nearly 8%.

Taken together, these results suggest that, consistent with our first hypothesis, the connectedness of the intrafirm inventor network structure is positively associated with a firm's ability to appropriate its technological knowledge, through channels that are not related to the type of innovation produced. We emphasize that while self-citation is a choice (i.e. inventors or managers choose what prior inventions to build upon and cite), our main dependent variable is a ratio which includes a component that is wholly outside of the firm's control: the number of citations made by its rivals. An important reason why we do not seek to establish a direct causal channel in our study is that the data generating process behind the ratio is likely to be iterative and complex. For example, the focal firm may initially cite its own prior invention, which being observable may alert a rival of its potential and lead to the rival attempting to build its own follow-on invention. Thus, we are interested in the net balance between self and others' citations, and especially in the timing of this race. We also note that, while prior work (e.g. Argyres et al. 2020) largely treated *Entropy* and *Giant* as alternative, substitutable measures of connectedness, the substantially higher magnitude of our results for *Negative Entropy* suggest that the two measures may capture distinct features of generative appropriability. Below we develop the point that the structure of the periphery components themselves (which is captured by *Negative Entropy* but not by *Giant*) is a hitherto unexplored dimension of organizational structure.

We note three other relationships in these estimations. First, the coefficient on *ln(Components)* is consistently negative ($p<0.014$). Thus, the greater the number of unconnected clusters of inventors in the firm's network, the lower the incidence of self-citation. As indicated above, the hypothesized association between *Negative Entropy* and *SelfRatio* is present even after controlling for the simple number of components. Second, both *Generality* and *Originality* are negatively associated with *SelfRatio*; this may indicate that patent families that integrate or generate knowledge that spans widely disparate technological domains will create more inter-firm spillovers, if only because the focal firm is unlikely to be able or willing to expand its boundaries into all of the fields that are affected by such innovations. Finally, *CitesToPublications* is positively associated with *SelfRatio*; we speculate that firms whose innovation is at the frontier of science may more effectively appropriate this knowledge to the extent that relatively few rivals have the absorptive capacity to quickly incorporate it into their innovative efforts.

Organizational structure and the dynamics of generative appropriability

Hypothesis 2 predicts that the positive relationship between network connectedness and appropriation of a focal innovation's knowledge will be strongest immediately after the focal invention's development, and will decline with time since the focal innovation. To investigate this question, and for ease of interpretation we first split the data into early and late citations, consistent with the view that the value of knowledge as a source of competitive advantage decays quickly over time. Thus, for our first set of analyses we construct variables *SelfRatioEarly* and *SelfRatioLate* to capture early and late citations, using 80 months as the cutoff point. This cutoff is based on the idea that patents on average take about 30 months to get granted (Hegde & Luo, 2018), so this period should capture the first waves of follow-on patenting.[9]

[INSERT TABLE 5 ABOUT HERE]

Table 5 presents results of estimations that explore this, coding early citations as those occurring within the first 80 months after initial application. Models 1-2 regress *SelfRatioEarly* on *Giant*, *Negative Entropy*, and the full set of controls from Table 5. Models 3-4 regress *SelfRatioLate* on the same independent variables. As Models 1-2 indicate, once we split the sample, *Negative Entropy* and *Giant* both have similar associations with *SelfRatioEarly* as they did with *SelfRatio*, but the coefficients have greater magnitude ($p<0.01$) . In these models, a one-standard-deviation increase in *Negative Entropy* and *Giant* are respectively associated with a 15% and an 8% increase in *SelfRatioEarly*. In contrast, when *SelfRatioLate* is the dependent variable (Models 3-4), the coefficients on *Negative Entropy* and *Giant* are an order of magnitude smaller ($p=0.953$ and $p=0.862$, respectively), suggesting that the associations are entirely driven by early citations.

[9] As detailed below, we also use an alternative continuous measure of timing.

This evidence is consistent with our second hypothesis: that any impact of network structure connectedness on appropriation is likely to be short-lived after a technology is made public by a patent application. However, considering that the early years are likely to be extremely valuable for an innovation, this brief window of protection is likely to be very valuable. We note a similar, albeit less pronounced, pattern for *CitesToPublications*. This variable is positively associated with both *SelfEarly* and *SelfLate*, but the magnitude of the coefficients falls to one-quarter to one-half the value in the *SelfLate* models vs. the *SelfEarly* models. While our results are robust to using different cut-offs in our definition of "early," such definitions may raise concerns about arbitrary cut-offs. Thus, to further explore the temporal dimension of generative appropriability, we generate the variable *Months_post*, which counts the number of months that have elapsed since the filing of the original priority application, and also the interaction term *Negative EntropyXmonth* or *GiantXmonth* as appropriate. The coefficient on the interaction term provides insight into temporal change in the connectedness-appropriability relationship and thus is central to our assessment of Hypothesis 2.

[INSERT TABLE 6 ABOUT HERE]

Table 6 presents results of the ensuing estimations. Models 1-2 present results for *Negative Entropy*. Model 1 introduces the main effect *Months_post*. The point estimate on *Months_post* is negative (p=0.000), indicating that the self-citation share of citations to a focal patent application declines over time; this is consistent with general expectations that a firm's knowledge eventually leaks out to other organizations. Model 2 introduces *Negative EntropyXmonths*. This variable is negative (p=0.000), indicating that the salutary effect of connectedness (as measured by *Negative Entropy*) on self-citation ratio declines over time. Put differently, the effect of *Negative Entropy* is strongest immediately after the focal patent application is filed, consistent with our prediction that network connectedness facilitates early gains for generative appropriability. With the addition of the interaction term, the coefficient on the main effect of *Negative Entropy* doubles in magnitude; now, a one-standard deviation increase in *Negative Entropy* at, say, the two-year mark after the seminal priority patent application date is associated with an increase in SelfRatio of nearly 20%, falling to 10% seven years after that application date. Models 3-4 present analogous estimations for *Giant*, with an identical pattern of results. For both *Negative Entropy* and *Giant*, the rate of decline in the benefit of connectedness is slightly less than 1% per month, and in both cases, if one takes the point estimates at face value, the connectedness-*SelfRatio* relationship is reduced to zero roughly ten years years after the seminal patent's application date (9¾ years for *Giant* and 11½ years for *Negative Entropy*).

These results are consistent with our second hypothesis: the positive relationship between network structure connectedness and appropriation is likely to be particularly acute in the early years after an innovation is made public by a patent application. Considering that the early years are likely to be extremely valuable for an innovation, due both to the ability to preempt others through follow-on innovation and to the likely obsolescence of an innovation's value over time, this brief window of protection is likely to be highly beneficial.

Post-hoc analysis: self-citation ratio and patent value

Finally, we examine the relationship between self-citations and patent value. As noted above, there is some (albeit inconsistent) evidence of this association, especially in a large cross-industry setting, suggesting that the relationship may be sensitive to sample selection (for example Hall, et al. 2005 found a positive association in firm-year value equations, and only for smaller firms). Therefore, it is important to support our assumption that self-citations reflect value-enhancing appropriability is valid within our chosen sample of large innovative firms, and at our patent-family level of observation. We estimate four models of the form:

$$Value_{jk} = \alpha + \beta_1 SelfRatio_{jk} + \gamma X_{kt-1} + \theta Z_{jk} + \delta Firm_k + \omega Year_t + \mu Tech_{jk} + \epsilon_{jk} \qquad (5)$$

where $Value_{jk}$ is each of the above-described value outcomes of interest for patent family j of firm k. Our independent variable $SelfRatio_{jk}$ is calculated for the entire life of the patent family; X is a vector of time-varying firm-specific covariates for firm k at year t-1; Z is a vector of patent-specific covariates averaged across all the applications in family j; and *Firm*, *Year* and *Tech* are fixed effects (the latter being the technology class for seminal patent application j). The coefficient of interest is β_1, which reflects the relationship between firm k's share of subsequent citations to patent family j and the private value realized by firm k from patent family j.

[INSERT TABLE 7 ABOUT HERE]

As indicated in Table 7, *SelfRatio* is positively associated with all four measures of patent value. These associations are economically significant: For example, given the point estimate in Model 1 ($p<0.005$), an increase in the self-citation ratio for a given patent family from the mean to one standard deviation above the mean is associated with an increase of almost seven months in the average renewal term of patents within that family; in other words, the firm pays to renew the patents for a longer period, suggesting that the firm expects to appropriate profit from these patents for longer time. This represents a 9% increase at the mean level of renewal. For the other measures, a one-standard-deviation increase in the self-citation ratio is associated with the firm paying to patent the technology in 0.25 more jurisdictions (Model 2, an 8% increase at the mean); with a 6% higher likelihood of being patented in the important triad of US, Japan, and Europe (Model 3), and with 0.3 more applications in the family (Model 4, an 8% increase at the mean). These results are consistent with other recent patent-level studies (Belenzon 2012; Wagner & Wakeman 2016) although less consistent with Hall et al. (2005).

One potential explanation for this difference in findings with Hall et al. (2005) is that we are able to include controls at the patent family level for originality and generality. Hall et al. (2005) noted that generality would be expected to affect the relationship between citations and value, and acknowledged that their inability to include it as control variable (because their regressions were at the firm-year level of analysis) could cause omitted variable bias in their results.

Two other relationships stand out in Table 4. First, the coefficient on *CitesToPublications* is consistently positive and of substantial magnitude across all four models (with p-values ranging from 0.002 to 0.010 depending on the model). This is consistent with received wisdom that patented innovations that are closer to cutting-edge science – and therefore presumably cite greater numbers of scientific publications – are likely to be more valuable to firms, all else equal. Second, the number of patent applications submitted by a firm in a given year is negatively associated with these measures of value; this may reflect different firms' patenting strategies.

In summary, we find that the self-citation ratio is indeed associated with private value at the patent-family level. While it may be argued that renewals and international filings are only reflective of a firm's willingness to spend a few thousand dollars, the literature supports the view that firms do not take these decisions lightly—otherwise all patents would be renewed to term and filed internationally.

Robustness checks

It is possible that the observed associations are artifacts of our choice of measures or of the model's functional form. To address this, we replicate our estimations using several alternative measures and specifications, the results of which appear in the online appendix. Although it has been customary to employ OLS estimation using the simple self-citation ratio as the dependent variable in prior studies, recent research has revealed potential biases and challenges to interpretation when the dependent variable is a ratio that is bounded to [0,1] (see Villadsen & Wulff 2019, 2020 for an overview and replication studies). We therefore replicate all specifications using a Fractional Response Model, following Papke and Wooldrige (1996).[10] We implement the

[10] Note to editor and reviewers: If desired, we can include an online appendix with robustness-test estimation results.

fractional models using the fracreg package in STATA 15.2, which maximizes ln$L$ according to the formula:

$$\ln\ L = \sum_{j=1}^{N} w_j\ y_j \ln\left\{\exp(x'_j\beta)/\{1+\exp(x'_j\beta)\}\right\} \tag{6}$$

where N is the sample size, y_j is the dependent variable, x_j are the covariates for patent family j. This specification fits a fractional response model for a dependent variable that is greater than or equal to 0 and less than or equal to 1. It uses a logit model for the conditional mean with a quasilikelihood estimator similar to the generalized linear models described in STATA's glm. Our results using fractional models are very similar to OLS, but we report OLS regression results given their more intuitive interpretation.

We also re-estimate all models using the count of self-citations as the dependent variable instead of self-citation ratio, while including the total number of citations as a regressor and re-estimate all models at the firm-year level. Our results are consistently qualitatively the same.

DISCUSSION AND CONCLUSION

Strategy scholars have generated considerable research on the relationship between inventor networks and innovation, as well as on the strategies firms use to appropriate the fruits of their innovation efforts. Less studied, however, is the relationship between these two important research streams. In this paper, we bridge this gap by theorizing that more connected internal organizational structures may increase appropriation through faster cumulative innovation. Empirically, we find that firms with more connected internal inventor networks are better able to build on their own knowledge relative to rivals' efforts to do so, and that this effect is particularly salient early in the lives of new patents, consistent with our argument that such structures help by facilitating faster follow-on innovations relative to their rivals. While this study does not seek to establish causality in these relationships, we are nonetheless able to rule out the most obvious alternative explanations:

that this association is simply driven by the types of patents produced by firms with different intra-organizational network structures, and that the observed relationship is simply a mechanical correlation between connectedness and self-citation. Secondarily, we provide large-scale evidence to clarify conflicting findings in the literature about whether self-citation-based measures are a good proxy for generative appropriability. Using several measures of patent value, we find evidence that higher forward self-citation ratios are positively associated with the private value that a firm obtains from a given patent family.

These findings are useful for managers because they suggest that, to the extent they can influence the structure of a firm's internal inventor network, they may be able to improve its innovative performance. Argyres et al. (2020) showed, for example, that centralizing a firm's formal budget R&D authority – a change that is well within the power of top managers – is associated with a subsequent increase in the connectedness of the firm's internal inventor network, and with a corresponding increase in the impact of the firm's innovations. On the other hand, they also showed that networks respond slowly, taking years to change, so it is clear that top managers cannot simply choose the optimal level of internal network connectedness. In view of this long lag between managerial intervention and network structural changes, it is likely that internal inventor networks are only weakly endogenous, however this is still an open question.

Relatedly, what are the determinants of an optimal level of connectedness for an internal inventor network? If greater connectedness is associated with a greater ability to appropriate returns to innovation, why do not all firms seek to maximize the level of such connectedness? A recent study suggests that such connectedness involves costs: Belderbos, Park and Carree (2020) found that firms with stronger internal linkages between its inventors suffer lower financial performance if they only conduct R&D in countries with strong intellectual property protection. The drivers of connectedness costs remain poorly understood, however. For example, some researchers may resist becoming more connected to others outside their immediate, requiring the firm to be more heavy-handed in making R&D assignments than is consistent with successful nurturing of innovation. Another is that firms that grow by acquisition face challenges of integrating networks of researchers across due to cross-national or intra-organizational cultural differences. A third possibility is that limiting connectedness may provide other kinds of innovation benefits to the firm that outweigh greater generative appropriability. One intriguing finding in our study is that entropy is roughly twice as strongly associated with self-citations than is relative giant component size. This is particularly interesting given that the two measures are so highly correlated (per Table 2, $\rho = -0.97$). As we have argued, the two measures capture innovation in the core, because connections to the giant component mechanically increase (decrease) giant and entropy. However, entropy captures additional nuances about the distribution of periphery components, because giant only measures the size of the main component relative to the periphery. Given the high degree of innovative activity that occurs outside of the central inventor network in most firms, future research should explore how innovation in the core of a firm's internal network differs from that in its periphery, and how the two interact.

Finally, future research should also study the interactions between network connectedness and other appropriation mechanisms studied in the literature, such as employee non-compete agreements, patent enforcement efforts, internal security arrangements, location choices, etc. One risk of greater network connectedness may be knowledge leakage, because more employees become aware of the firm's recent innovations. Do firms featuring greater connectedness employ more security measures, or use more non-compete agreements, to reduce this risk? Do firms located in highly innovative geographic clusters such as Silicon Valley, where knowledge spillovers are more frequent, feature less network connectedness? Or are such firms able to use complementary mechanisms to mitigate the risk of leakage enough such that locating in such regions as a highly connected firm poses little risk relative to the benefits of colocation? For example, do more connected firms defend their patents more aggressively? There is clearly much to be learned about how firms appropriate returns from their innovation efforts.

REFERENCES

Abrams, D., Akcigit, U., & Grennan, J. (2018). Patent value and citations: Creative destruction of strategic disruption? *NBER working paper.*

Agarwal, R., Ganco, M., & Ziedonis, R. H. (2009). Reputations for toughness in patent enforcement: Implications for knowledge spillovers via inventor mobility. *Strategic Management Journal*, 30(13): 1349–1374.

Ahuja, G., Lampert, C. M., & Novelli, E. (2013). The second face of appropriability: Generative appropriability and its determinants. *Academy of Management Review*, *38*(2): 248–269.

Alcácer, J., & Chung, W. (2007). Location strategies and knowledge spillovers. *Management Science*, 53(5): 760–776.

Amburgey, T. L., Al-Laham, A., Tzabbar, D., & Aharonson, B. (2008). The structural evolution of multiplex organizational networks: research and commerce in biotechnology. In J. A. C. Baum & T. J. Rowley (Eds.), *Network Strategy (Advances in Strategic Management)* 25: 171-209). Emerald Group Publishing Limited.

Argyres, N., Rios, L. A., & Silverman, B. S. (2020). Organizational change and the dynamics of innovation: Formal R&D structure and intrafirm inventor networks. *Strategic Management Journal*, 41(11): 2015-2049.

Arora, A., Belenzon, S., & Rios, L. (2014). Make, buy, organize: The interplay between research, external knowledge, and firm structure. *Strategic Management Journal*, 35(3): 317-337.

Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors* (pp. 609–626). *NBER.*

Balachandron, S. & Eklund, J. (2019). Innovation via interfirm collaboration: The role of partner organization design. working paper, Marshall School of Business, USC

Barney, J. B. (1986). Strategic factor markets: Expectations, luck, and business strategy. *Management Science*, 32(10): 1231--1241.

Baum, J. A. C., Shipilov, A. V., & Rowley, T. J. (2003). Where do small worlds come from? *Industrial and Corporate Change*, 12(4): 697--725.

Belderbos, R., Park, J. & Carree, M. (2020). Do R&D investments in weak IPR countries destroy market value? The role of internal linkages. *Strategic Management Journal*, 42(8): 1401-1431.

Birkinshaw, J. & Hood, M. (1998). Multinational subsidiary evolution: Capability and charter change in foregn-owned subsidiary companies. *Academy of Management Review*, 23(4): 773-795.

Bloom, N., Schankerman, M.,& Van Reenan, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81: 1347-1393.

Bollobas, B. 2001. *Random Graphs*, Cambridge Studies in Advanced Mathematics (Book 73), 2nd edition, Cambridge University Press, Caombridge.

Cannon, M. & Edmondson, A. (2005). Failing to learn and learning to fail (intelligently): How great organizations put failure to work to innovate and improve. *Long Range Planning*, 38(3): 299-

319.
Carnanbuci, G., & Operti, E. (2013). Where do firms' recombinant capabilities come from? Intraorganizational networks, knowdege, and firms' ability to innovate through technological recombination. *Strategic Management Journal*, 34: 1591–1613.
Ceccagnoli, M. (2005). Firm heterogeneity, imitation, and the incentives for cost reducing R&D effort. *Journal of Industrial Economics*, *53*(1): 83–100.
Cohen, W. M., & Klepper, S. (1996). A reprise of size and R & D. *The Economic Journal*, 106(437): 925–951.
Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2000). *Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)* (NBER WORKING PAPER SERIES No. 7552). National Bureau of Economic Research Cambridge, Mass., USA.
Dosi, G. (1988). Sources, procedures, and Mmcroeconomic effects of innovation. *Journal of Economic Literature*, 26(3): 1120–1171.
Eggers, J.P. (2012). All experience is not equal: Learning, adapting and focusing in product portfolio management. *Strategic Management Journal*, 33(3): 315-335.
Furman, J. L., & Stern, S. (2006). Climbing atop the shoulders of giants: The impact of institutions on cummulative research. *American Economic Review*, 101(5): 1933-1963.
Galasso, A., & Schankerman, M. (2014). Patents and cumulative innovation : causal evidence from the courts. *NBER Working Papers Series*.
Ghemawat, P. (1987). Investment in lumpy capacity. *Journal of Economic Behavior & Organization,* 8(2): 265—277.
Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, 17(10): 109–122.
Grigoriou, K., & Rothaermel, F. T. (2017). Organizing for knowledge generation: Internal knowledge networks and the contingent effect of external knowledge sourcing. *Strategic Management Journal*, 38(5): 961–964.
Guimera, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308(5722), 697-702.
Guler, I., & Nerkar, A. (2012). The impact of global and local cohesion on innovation in the pharmaceutical industry. *Strategic Management Journal*, 33: 535–549.
Khanna, R., Guler, I., Nerkar, A. 2016. Fail often, fail big and fail fast? Learning from small failures and R&D performance in the pharmaceutical industry. *Strategic Management Journal*, 59(2): 436-459.
Kuhn, J., Younge, K., & Marco, A. (2020). Patent citations reexamined. *RAND Journal of Economics*, 51(1): 109-132.
Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, 36(1): 16–38.
Hall, B.H. & Harhoff, D. (2012). Recent research on the economics of patents. *NBER Working Paper.*
Hegde, Deepak, and Hong Luo. "Patent publication and the market for ideas." *Management Science* 64.2 (2018): 652-672.
Jaffe, A. B., Trajtenberg, M., & Henderson, R. M. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3): 577–598.
Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3(3): 383–397.
Kogut, B., & Zander, U. (1993). Knowledge of the firm and the evolutionary theory of the

multinational corporation. *Journal of International Business Studies*, 24(4): 625--645.
Kogut, B., & Zander, U. (1996). What Firms Do? Coordination, Identity, and Learning. *Organization Science*, 7(5): 502–518.
Lauerman, J. (2022). Merck's COVID pill fumble gives Pfizer potential 417B win. Bloomberg Businessweek. February 3 from https://www.bloomberg.com/news/articles/2022-02-03/drug-maker-merck-s-antiviral-covid-pill-fumble-gives-pfizer-17-billion-win.
Lieberman, M. B., & Montgomery., D. B. (1988). First-mover advantages. *Strategic Management Journal*, 9: 41-58.
Liebeskind, J. P. (1996). Knowledge, strategy, and the theory of the firm. *Strategic Management Journal*, 17(S2): 93-107.
Mansfield, E. (1961). Technical change and the rate of imitation. *Econometrica*, 29(4): 741–766.
Mansfield, E., Schwartz, M., & Wagner, S. (1981). Imitation costs and patents : An empirical study. *The Economic Journal*, 91(364): 907–918.
Martínez, C. (2011). Patent families: When do different definitions really matter? *Scientometrics*, 86(1), 39–63.
Marx, M., Strumsky, D., & Fleming, L. (2009). Mobility, skills, and the Michigan non-compete experiment. *Management Science*, 55(6): 875–889.
Mitchell, W. (1991). Dual clocks: Entry order influences on incumbent and newcomer market share and survival when specialized assets retain their value. *Strategic Management Journal*, 12(2): 85-100.
Moreira, S., Markus, A., & Laursen, K. (2018). Knowledge diversity and coordination: The effect of intrafirm inventor task networks on absorption speed. *Strategic Management Journal*, 39(9): 2517–2546.
Moser, P., Ohmstedt, J., & Rhode, P. W. (2018). Patent citations-an analysis of quality differences and citing practices in hybrid corn. *Management Science*, 64(4): 1926–1940.
Nahapiet, J., & Ghoshal, S. (1998). Social capital , intellectual capital , and the organizational advantage. *Academy of Management Review*, 23(2): 242–266.
Nelson, R. & Winter, S. 1982. *An Evolutionary Theory of Economic Change*, Cambridge, MA: Belknap Press.
Obstfeld, D. (2005). Social networks, the tertius iungens orientation, and involvement in innovation. *Administrative Science Quarterly*, 50(1): 100-130.
Pakes, A. (1986). Patents as options: Some estimates of the value of holding European patent stocks. *Econometrica*, 54: 755-784.
Rajan, R. G., & Zingales, L. (2001). The firm as a dedicated hierarchy: A theory of the origins and growth of firms. *Quarterly Journal of Economics*, 116(3): 805–851.
Reagans, R., & McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative Science Quarterly*, 48(2): 240–267.
Reagans, R., & Zuckerman, E. W. (2001). Networks, diversity, and productivity: The social capital of corporate R&D teams. *Organization Science*, 12(4): 502-517.
Rios, L. A. (2021). On the origin of technological acquisition strategy: The interaction between organizational plasticity and environmental munificence. *Strategic Management Journal,* 42:1299–1325.
Rivkin, J. (2001). Reproducing knowledge: Replication without imitation at moderate complexity. *Organization Science*, 12(3): 274-293.
Rosenberg, N. (1976). *Perspectives on technology*. New York: Cambridge University Press.
Rothaermel, F. (2001). Incumbent's advantage through exploiting complementary assets via interfirm cooperation. *Strategic Management Journal*, 22(6-7): 687-699.
Schankerman, M., & Pakes, A. (1986). Estimates of the value of patent rights in European

countries during the post-1950 period. *Economic Journal*, 96(384): 1052-1076.
Schilling, M. A. (2015). Technology shocks, technological collaboration, and innovation outcomes. *Organization Science*, 26(3): 668–686.
Schilling, M. A., & Phelps, C. C. (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. *Management Science*, 53(7): 1113–1126.
Scotchmer, S. (1991). Standing on the shoulders of giants: cumulative research and the patent law. *The Journal of Economic Perspectives*, *5*(1), 29-41.
Serrano, C. J. (2018). Estimating the gains from trade in the market for patent rights. *International Economic Review*, *59*(4), 1877–1904.
Sharapov, D. & Macaulay, S. (2022). Design as an isolating mechanism for capturing value from innovation: From cloaks and traps to sabotage. *Academy of Management Review*, 47(1): 139-161.
Singh, J., & Agrawal, A. (2011). Recruiting for ideas: How firms exploit the prior inventions of new hires. *Management Science*, 57(1), 129-150.
Singh, J., & Marx, M. (2013). Geographic constraints on knowledge spillovers: Political borders vs. spatial proximity. *Management Science*, 59(9): 2056-2078.
Sitkin, S. (1992). Learning through failure: The strategy of small losses. *Research in Organizational Behavior,* 14: 231-267.
Somaya, D. (2012). Patent strategy and management: An integrative review and research agenda. *Journal of Management*, 38(4): 1084-1114.
Spender, J.C. (1996). Making knowledge the basis of a dynamic theory of the firm. *Strategic Management Journal*, 17: 45-62.
Sorenson, O., Rivkin, J. W., & Fleming, L. (2006). Complexity, networks and knowledge flow. *Research Policy*, 35(7): 994--1017.
Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, 15(6): 285–305.
Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7): 509–533.
Tripsas, M. (1997). Unravelling the process of creative destruction: Complementary assets and incumbent survival in the typesetter industry. *Strategic Management Journal*, 18(S1): 119-142.
Tzabbar, D., Silverman, B. S., & Aharonson, B. S. (2015). Learning by hiring or hiring to avoid learning? *Journal of Managerial Psychology*, 30(5): 550–564.
Winter, S. (1987). Knowledge and competences as strategic assets. In Teece, D.J. (Ed.), *The Competitive Challenge: Strategies for Industrial Innovation and Renewal* (pp. 159-184). Cambridge, MA: Ballinger.
Winter, S. & Szulanaki, G. (2001). Replication as strategy. *Organization Science*, 12(6): 730-743.
Zhao, M. (2006). Conducting R&D in countries with weak intellectual property rights protection. *Management Science*, 52(8): 1185–1199.

Figure 1: Example of domestic patent family.

Square, Inc's initial application for its payment technology was ultimately divided into dozens of continuations and continuations-in-part which resulted in nine different granted patents covering the same underlying innovation, and claiming the same priority date. These "priority" or "continuation" families are distinct from Jurisdiction families (which are equivalent patents in different countries). A key feature of continuation families is that all the applications must have the same technical content (drawings and descriptions) while allowing for updated claims (often to keep up with rivals' advances). Thus, when considering the share of forward citations generated by the original inventing firm relative to all other entities, it is important to aggregate to the family level in order to properly capture the impact of a technology. For the most valuable inventions, it is not unusual to see complicated family trees such as the one below from Square.

(75) Inventors: **Amish Babu**, San Francisco, CA (US); **Gregory Staples**, Petaluma, CA (US); **Jack Dorsey**, San Francisco, CA (US); **James M. McKelvey**, Miami, FL (US)

Related U.S. Application Data

is a division of application No. 13/179,836, filed on Jul. 11, 2011, now abandoned, which is a continuation-in-part of application No. 12/707,228, filed on Feb. 17, 2010, now abandoned, which is a continuation of application No. 12/050,752, filed on Mar. 18, 2008, now Pat. No. 7,684,809, which is a continuation of application No. 10/355,557, filed on Jan. 31, 2003, now Pat. No. 7,376,431, said application No. 13/179,836 is a continuation-in-part of application No. 12/903,753, filed on Oct. 13, 2010, and a continuation-in-part of application No. 12/903,823, filed on Oct. 13, 2010, now Pat. No. 8,534,546, and a continuation-in-part of application No. 13/005,822, filed on Jan. 13, 2011, now Pat. No. 8,870,070, and a continuation-in-part of application No. 12/985,982, filed on Jan. 6, 2011, now Pat. No. 8,573,486, and a continuation-in-part of application No. 13/010,976, filed on Jan. 21, 2011, and a continuation-in-part of application No. 13/012,495, filed on Jan. 24, 2011, now Pat. No. 8,500,018, and a continuation-in-part of application No. 13/043,203, filed on Mar. 8, 2011, now Pat. No. 8,573,487, and a continuation-in-part of application No. 13/043,258, filed on Mar. 8, 2011, now Pat. No. 8,870,071, and a continuation-in-part of application No. 13/043,263, filed on Mar. 8, 2011, now Pat. No. 8,876,003, and a continuation-in-part of application No. 13/043,268, filed on Mar. 8, 2011, now Pat. No. 8,302,860, and a continuation-in-part of application No. 13/043,270, filed on Mar. 8, 2011, now Pat. No. 8,235,287.

(60) Provisional application No. 60/355,667, filed on Feb. 5, 2002, provisional application No. 60/356,861, filed on Feb. 12, 2002, provisional application No. 60/361,646, filed on Mar. 4, 2002.

Figure 2: Example of citations to application prior to patent grant

Top panel shows the granted patent, published on April 13, 2010. Bottom panel shows patent application, published on March 9, 2006. The application received 97 citations prior to the grant date. The granted patent has received 42 citations through June, 2022. This technology was bought by Google and crucial to the development of Google Glass, and was cited by many rivals competing to enter that niche, such as Microsoft. Note that application predates purchase by Google, which is not named on its face. This illustrates the importance of aggregating citations to related patents to capture generative appropriability.

(12) **United States Patent**
Adam et al.

(10) **Patent No.: US 7,697,735 B2**
(45) **Date of Patent: Apr. 13, 2010**

(54) **IMAGE BASED MULTI-BIOMETRIC SYSTEM AND METHOD**

(75) Inventors: **Hartwig Adam**, Beverly Hills, CA (US); **Hartmut Neven**, Malibu, CA (US); **Johannes B. Steffens**, Westoverledingen (DE)

(73) Assignee: **Google Inc.**, Mountain View, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 871 days.

(21) Appl. No.: **11/158,906**

(22) Filed: **Jun. 21, 2005**

(65) **Prior Publication Data**

US 2006/0050933 A1 Mar. 9, 2006

Related U.S. Application Data

(60) Provisional application No. 60/581,496, filed on Jun. 21, 2004.

6,944,318	B1 *	9/2005	Takata et al.	382/115
7,415,138	B2 *	8/2008	Schneider et al.	382/115
7,440,929	B2 *	10/2008	Schneider et al.	706/15
2002/0136435	A1 *	9/2002	Prokoski	382/118
2004/0052418	A1	3/2004	DeLean	
2004/0252865	A1 *	12/2004	Tisse	382/117
2005/0185060	A1	8/2005	Neven	
2006/0012677	A1	1/2006	Neven et al.	
2006/0050933	A1 *	3/2006	Adam et al.	382/118
2006/0093183	A1 *	5/2006	Hosoi	382/103
2007/0253604	A1 *	11/2007	Inoue et al.	382/118

(Continued)

FOREIGN PATENT DOCUMENTS

DE 10245900 A1 4/2004

(Continued)

(19) **United States**
(12) **Patent Application Publication**
Adam et al.

(10) **Pub. No.: US 2006/0050933 A1**
(43) **Pub. Date: Mar. 9, 2006**

(54) **SINGLE IMAGE BASED MULTI-BIOMETRIC SYSTEM AND METHOD**

(76) Inventors: **Hartwig Adam**, Beverly Hills, CA (US); **Hartmut Neven**, Malibu, CA (US); **Johannes B. Steffens**, Westoverledingen (DE)

Correspondence Address:
ROBROY R FAWCETT
1576 KATELLA WAY
ESCONDIDO, CA 92027 (US)

(21) Appl. No.: **11/158,906**

(22) Filed: **Jun. 21, 2005**

Related U.S. Application Data

(60) Provisional application No. 60/581,496, filed on Jun. 21, 2004.

Publication Classification

(51) **Int. Cl.**
G06K 9/00 (2006.01)

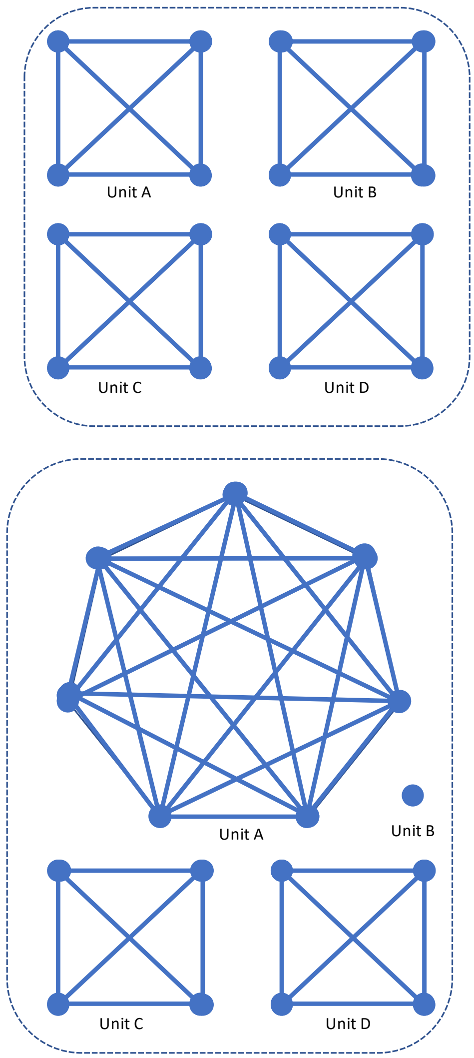
(52) **U.S. Cl.** ... **382/118**

(57) **ABSTRACT**

This disclosure describes methods to integrate face, skin and iris recognition to provide a biometric system with unprecedented level of accuracy for identifying individuals. A compelling feature of this approach is that it only requires a single digital image depicting a human face as source data.

Figure 3 Illustration of Metcalfe's Law

As inventors migrate to larger components, there is a positive non-linear increase in the upper limit of *possible* number of connections among them (for any fixed set of components and inventors) In Panel A, four units with four inventors each can only achieve a maximum of six inventor ties per unit (24 at the firm level). However, moving three inventors from Unit B to Unit A increases the maximum number of connections to 21 for Unit A (an increase of 15) while only reducing the number of connections for Unit B by six (down to zero), bringing the total for the firm to 34 connections (note that Unit A could be any component, either the giant or a peripheral one).



Normalized Entropy=1
4 Units
16 Inventors
4,4,4,4 Inventors per unit
6 Maximum knowledge recombinations at unit level
24 total recombinations at firm level

Normalized Entropy=0.89
4 Units
16 Inventors
7,4,4,1 Inventors per unit
21 Maximum knowledge recombinations at unit level
34 total recombinations at firm level

Table 1: Variable definitions

$SelfRatio_{jk}$	(count of all citations by firm k to patents or patent applications in firm k's patent family j through 2019) / (count of all citations to patents or patent applications in firm k's patent family j through 2019)
$SelfRatioEarly_{jk}$	(count of all citations by firm k to patents or patent applications in firm k's patent family j during the first 80 months after the date of application for the seminal patent application in patent family j) / (count of all citations to patents or patent applications in firm k's patent family j during the first 80 months after the date of application for the seminal patent application in patent family j)
$SelfRatioLate_{jk}$	(count of all citations by firm k to patents or patent applications in firm k's patent family j more than 80 months after the date of application for the seminal patent application in firm k's patent family j and through 2019) / (count of all citations to patents or patent applications in firm k's patent family j more than 80 months after the date of application for the seminal patent application in firm k's patent family j and through 2019)
$SelfRatio\text{-}HJT_{ik}$	(count of all citations by firm k to firm k's patent i through 2019) / (count of all citations to firm k's patent i through 2019), per Hall et al. (2005).
$Giant_{kt}$	(count of inventors in the largest component of firm k's inventor network in year t) / (count of all inventors in firm k's inventor network in year t)
$Entropy_{kt}$	A variant of 1 - herfindahl index of the sizes of components of firm k's inventor network in year t. See equations (1) and (2) in the text for precise definition.
$Generality_{jk}$	1 - herfindahl index of the technology classes of patent applications that cite firm k's patent family j, per Hall et al. (2001)
$Originality_{jk}$	1 - herfindahl index of the technology classes of patent applications that are cited by firm k's patent family j, per Hall et al. (2001)
$CitesToPublications_{jk}$	Count of the number of citations to scientific publications that appear on patent applications in firm k's patent family j
$Ln(Family\ Age)_{jk}$	Number of years between application year of seminal application and application year for the most recent application in firm k's patent family j, as of 2013
$Ln(Patents)_{kt-1}$	Natural log of (1 + the number of patent applications submitted by firm k) in year t-1
$Ln(Components)_{kt-1}$	Natural log of (1 + the number of distinct components in firm k's inventor network) in year t-1
$Ln(R\&D)_{kt-1}$	Natural log of (1 + firm k's R&D expenditure) in year t-1
$Ln(Assets)_{kt-1}$	Natural log of (1 + firm k's assets) in year t-1
$Ln(Employees)_{kt-1}$	Natural log of (1 + firm k's employees) in year t-1
$Ln(Sales)_{kt-1}$	Natural log of (1 + firm k's revenue) in year t-1
$YrsRenewed_{jk}$	Average number of years that the granted patents in firm k's patent family j are renewed, through 2013
$Jurisdictions_{jk}$	Count of the number of countries in which at least one application in firm k's patent family j is submitted
$Triad_{jk}$	Set equal to 1 if at least one application in firm k's patent family j is granted in all three of these jurisdictions: US, Japan, and Europe; set equal to 0 otherwise
Patent family $size_{jk}$	Count of the total number of patent applications in firm k's patent family j as of 2013
Patent $complexity_{jk}$	the Fleming/Sorenson complexity measure averaged for all applications in firm k's seminal patent family j

Table 2: Summary Statistics

Patent-Family unit of analysis

	N	Mean	Std Dev	Min	Max
Giant Component $_{kt-1}$	430,977	0.420	0.216	0.015	1
Negative Entropy $_{kt-1}$	430,977	0.491	0.164	-0.940	0
SelfRatio$_{jk}$	462,202	0.163	0.277	0	1
SelfRatioEarly$_{jk}$	317,471	0.235	0.338	0	1
SelfRatioLate$_{jk}$	332,826	0.115	0.239	0	1
SelfRatio-HJT$_{jk}$	224,574	0.145	0.276	0	1
Jurisdictions$_{jk}$	408,122	1.864	3.190	1	45
PatentFamilySize$_{jk}$	408,122	3.946	4.384	1	298
Triad$_{jk}$	408,122	0.448	0.497	0	1
YrsRenewed$_{jk}$	408,122	6.002	5.896	0	25
Generality$_{jk}$	430,977	0.964	0.530	0	3.428
Originality$_{jk}$	430,977	1.008	0.254	0	1.757
CitesToPublications$_{jk}$	430,977	0.947	3.138	0	220.616
ln(Family Age) $_{jk}$	430,977	0.349	0.669	0	3.367
ln(Patents)$_{kt-1}$	430,977	6.225	1.652	0.693	9.234
ln(Components)$_{kt-1}$	430,977	5.406	1.346	0.693	7.283
ln(R&D)$_{kt-1}$	430,977	6.806	1.692	0	9.407
ln(Assets)$_{kt-1}$	430,977	9.731	1.714	1.253	13.081
ln(Employees)$_{kt-1}$	430,977	3.919	1.427	0.005	6.777
ln(Sales)$_{kt-1}$	430,977	9.509	1.748	0	12.449

Firm-Year unit of analysis

	N	Mean	Std Dev	Min	Max
Giant Component $_{kt-1}$	430,977	0.392	0.216	0.015	1
Entropy $_{kt-1}$	430,977	0.515	0.164	0	0.940
Components$_{kt-1}$	430,977	52.77	1.346	0.693	7.283
R&D$_{kt-1}$	430,977	822	1.692	0	9.407
Assets$_{kt-1}$	430,977	6,931	1.714	1.253	13.081
Employees$_{kt-1}$	430,977	47.43	1.427	0.005	6.777
Sales$_{kt-1}$	430,977	16,475	1.748	0	12.449

Table 3: Correlation Table (patent-family unit of analysis)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 Giant component	1.00																	
2 Entropy	-0.97	1.00																
3 SelfRatioEarly	0.10	-0.11	1.00															
4 SelfRatioLate	0.04	-0.04	0.40	1.00														
5 Jurisdictions	0.00	-0.03	0.08	0.07	1.00													
6 PatentFamilySize	-0.02	-0.02	0.09	0.07	0.76	1.00												
7 Triad	0.00	-0.02	0.06	0.05	0.68	0.68	1.00											
8 YrsRenewed	-0.13	0.16	0.00	0.03	-0.29	-0.34	-0.33	1.00										
9 Generality	0.00	-0.01	-0.05	-0.04	0.05	0.08	0.06	-0.01	1.00									
10 Originality	0.01	-0.01	-0.01	-0.03	0.04	0.06	0.04	-0.02	0.31	1.00								
11 CitesToPublications	0.01	-0.01	0.05	0.03	0.04	0.09	0.03	0.03	0.06	0.07	1.00							
12 ln(Family Age)	-0.01	-0.01	0.08	0.07	0.34	0.57	0.28	-0.09	0.08	0.06	0.11	1.00						
13 ln(Patents)	0.41	-0.40	0.06	0.03	-0.14	-0.22	-0.12	-0.00	-0.07	-0.03	-0.04	-0.13	1.00					
14 ln(Components)	0.11	-0.08	0.02	0.01	-0.11	-0.20	-0.09	0.06	-0.06	-0.03	-0.04	-0.12	0.84	1.00				
15 ln(R&D)	0.31	-0.30	0.02	-0.01	-0.10	-0.16	-0.10	-0.02	-0.05	-0.02	-0.03	-0.10	0.84	0.78	1.00			
16 ln(Assets)	0.32	-0.31	0.04	0.01	-0.09	-0.14	-0.07	-0.07	-0.03	-0.01	-0.04	-0.13	0.76	0.72	0.87	1.00		
17 ln(Employees)	0.17	-0.14	0.03	0.01	-0.11	-0.18	-0.09	-0.01	-0.04	-0.00	-0.04	-0.14	0.72	0.75	0.80	0.91	1.00	
18 ln(Sales)	0.25	-0.24	0.03	0.01	-0.11	-0.17	-0.09	-0.05	-0.03	-0.01	-0.04	-0.14	0.75	0.73	0.86	0.97	0.94	1.00

Table 4: Self-citation ratio as a function of inventor network structure, patent family level (*SelfRatio*)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Negative Entropy	0.125	0.124	0.115	0.115				
	(0.000)	(0.000)	(0.000)	(0.000)				
Giant					0.057	0.059	0.054	0.053
					(0.003)	(0.003)	(0.003)	(0.003)
ln(Patents)	-0.002	-0.002	-0.002	-0.002	0.002	0.002	0.003	0.000
	(0.612)	(0.674)	(0.336)	(0.334)	(0.486)	(0.427)	(0.904)	(0.907)
ln(Components)	-0.025	-0.024	-0.017	-0.018	-0.030	-0.030	-0.024	-0.023
	(0.002)	(0.002)	(0.014)	(0.014)	(0.000)	(0.000)	(0.001)	(0.001)
ln(R&D)	0.017	0.018	0.015	0.015	0.018	0.017	0.015	0.015
	(0.014)	(0.014)	(0.029)	(0.029)	(0.019)	(0.020)	(0.037)	(0.037)
ln(Assets)	0.013	0.014	0.009	0.009	0.015	0.015	0.011	0.011
	(0.082)	(0.084)	(0.167)	(0.168)	(0.030)	(0.031)	(0.081)	(0.082)
ln(Employees)	-0.036	-0.036	-0.032	-0.032	-0.036	-0.036	-0.032	-0.032
	(0.038)	(0.035)	(0.064)	(0.064)	(0.047)	(0.044)	(0.076)	(0.076)
ln(Sales)	0.015	0.014	0.011	0.011	0.016	0.014	0.012	0.012
	(0.027)	(0.046)	(0.115)	(0.118)	(0.027)	(0.044)	(0.113)	(0.117)
Originality		-0.025	-0.009	-0.010		-0.025	-0.009	-0.010
		(0.001)	(0.097)	(0.070)		(0.001)	(0.096)	(0.069)
CitesToPublications		0.003	0.003	0.003		0.003	0.003	0.003
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Generality			-0.026	-0.026			-0.027	-0.027
			(0.002)	(0.003)			(0.002)	(0.003)
ln(Family Age)			0.010	0.010			0.010	0.010
			(0.000)	(0.000)			(0.000)	(0.000)
Complexity				-0.000				-0.000
				(0.714)				(0.713)
Constant	0.137	0.161	0.203	0.202	0.034	0.062	0.117	0.112
	(0.091)	(0.046)	(0.005)	(0.005)	(0.567)	(0.314)	(0.047)	(0.046)
Observations	430,977	423,023	400,478	400,462	430,977	423,023	400,478	400,462
Adjusted R^2	0.15	0.15	0.17	0.17	0.15	0.15	0.17	0.17
Year, Firm, Tech FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimation; dependent variable is SelfRatio; p-values in parentheses.

Table 5:

Self-citation ratio as a function of inventor network structure
early citations (*SelfCiteEarly*) vs. later citations (*SelfCiteLate*)

	SelfRatioEarly		*SelfRatioLate*	
	(1)	(2)	(3)	(4)
Negative Entropy	0.163		0.001	
	(0.003)		(0.953)	
Giant component		0.080		-0.002
		(0.013)		(0.862)
ln(Patents)	-0.005	0.000	-0.005	-0.004
	(0.407)	(0.926)	(0.372)	(0.405)
ln(Components)	0.011	0.003	-0.003	-0.003
	(0.227)	(0.747)	(0.659)	(0.568)
ln(R&D) t-1	0.002	0.002	-0.002	-0.002
	(0.861)	(0.851)	(0.756)	(0.749)
ln(Assets) t-1	-0.002	-0.003	0.003	0.003
	(0.861)	(0.836)	(0.702)	(0.706)
ln(Emp) t-1	-0.009	-0.009	-0.017	-0.017
	(0.695)	(0.703)	(0.222)	(0.224)
ln(Sales) t-1	-0.007	-0.006	0.001	0.001
	(0.667)	(0.705)	(0.848)	(0.841)
Originality	0.001	0.001	-0.019	-0.019
	(0.818)	(0.809)	(0.000)	(0.000)
CitesToPublications	0.004	0.004	0.001	0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Generality	-0.042	-0.042	-0.019	-0.019
	(0.000)	(0.000)	(0.004)	(0.004)
ln(Family Age)	0.030	0.030	0.013	0.014
	(0.000)	(0.000)	(0.000)	(0.000)
Complexity	-0.012	-0.096	-0.005	-0.010
	(0.876)	(0.729)	(0.675)	(0.507)
Constant	0.438	0.326	0.228	0.232
	(0.002)	(0.009)	(0.005)	(0.003)
Observations	317,471	317,471	332,826	332,826
Adjusted R^2	0.15	0.15	0.16	0.16
Year, Firm & Tech FE	Yes	Yes	Yes	Yes

Notes: OLS estimation; p-values in parentheses

Table 6: Self-citation ratio as function of inventor network structure and time since focal innovation:

	(1)	(2)	(3)	(4)
Negative Entropy	0.094	0.221		
	(0.004)	(0.000)		
Giant			0.055	0.14
			(0.009)	(0.000)
Entropy x month		-0.0016		
		(0.000)		
Giant x month				-0.001
				(0.000)
Months post	-0.001	-0.001	-0.001	-0.000
	(0.000)	(0.000)	(0.000)	(0.084)
Generality	-0.028	-0.028	-0.028	-0.028
	(0.001)	(0.001)	(0.001)	(0.001)
Originality	-0.0094	-0.010	-0.0093	-0.010
	(0.109)	(0.091)	(0.110)	(0.095)
NPL Cites	0.0028	0.0028	0.0028	0.0028
	(0.000)	(0.000)	(0.000)	(0.000)
ln(famage)	0.017	0.019	0.018	0.019
	(0.000)	(0.000)	(0.000)	(0.000)
Patents/year t-1	-0.0021	-0.005	0.000	-0.002
	(0.689)	(0.402)	(0.967)	(0.779)
ln(components)	-0.0034	0.003	-0.007	-0.003
	(0.611)	(0.713)	(0.310)	(0.708)
ln(R&D) t-1	0.00075	0.000	0.001	0.001
	(0.931)	(0.984)	(0.914)	(0.930)
ln(Assets) t-1	-0.0023	0.000	-0.002	0.001
	(0.810)	(0.956)	(0.833)	(0.983)
ln(Emp) t-1	-0.011	-0.016	-0.012	-0.016
	(0.516)	(0.322)	(0.501)	(0.330)
ln(Sales) t-1	-0.0022	-0.002	-0.002	-0.012
	(0.844)	(0.888)	(0.869)	(0.931)
Constant	0.401	0.440	0.332	0.228
	(0.000)	(0.000)	(0.000)	(0.001)
Observations	400,282	400,282	400,282	400,282
R^2	0.19	0.19	0.19	0.19
Adjusted R^2	0.19	0.19	0.19	0.19
Year Firm Tech FE	Yes	Yes	Yes	Yes

Notes: OLS estimation. Dependent variable is SelfRatio; p-values in parentheses.

Table 7: Patent-family value as a function of self-citation ratio

	(1) YrsRenewed	(2) Jurisdictions	(3) Triad	(4) Patent Family Size
SelfRatio	2.017	0.910	0.095	1.101
	(0.005)	(0.000)	(0.000)	(0.000)
ln(Patents)	-0.537	-0.183	-0.024	-0.261
	(0.008)	(0.047)	(0.347)	(0.003)
ln(Components)	0.123	0.445	0.069	0.261
	(0.692)	(0.000)	(0.004)	(0.023)
ln(R&D)	-0.671	0.094	0.060	0.152
	(0.037)	(0.433)	(0.052)	(0.247)
ln(Assets)	-0.771	-0.069	0.021	-0.125
	(0.027)	(0.534)	(0.476)	(0.330)
ln(Employees)	1.643	-0.396	-0.132	-0.366
	(0.000)	(0.001)	(0.001)	(0.007)
ln(Sales)	-0.683	0.237	0.051	0.303
	(0.080)	(0.056)	(0.093)	(0.026)
Originality	-0.114	0.014	0.006	0.048
	(0.120)	(0.669)	(0.385)	(0.228)
CitesToPublications	0.062	0.012	0.002	0.021
	(0.000)	(0.001)	(0.010)	(0.000)
Generality	0.058	0.021	0.006	0.024
	(0.143)	(0.188)	(0.189)	(0.111)
ln(Family Age)	-1.053	1.183	0.193	2.073
	(0.000)	(0.000)	(0.000)	(0.000)
Complexity	-0.035	-0.052	0.617	0.078
	(0.185)	(0.352)	(0.814)	(0.661)
Constant	20.762	0.568	-0.464	1.503
	(0.000)	(0.548)	(0.002)	(0.180)
Observations	408122	408122	408122	408122
Adjusted R^2	0.390	0.399	0.270	0.197
Year, Firm, & Tech FE.	Yes	Yes	Yes	Yes

Notes: OLS estimation; p-values in parentheses.