



---

Faculty Publications

---

2016

## Pipes, Pools and Filters: How Collaboration Networks Affect Innovative Performance

Harpeet Singh  
*University of Texas at Dallas*

David Kryscynski  
*Brigham Young University, dk@byu.edu*

Xinxin Li  
*University Of Connecticut*

Ram Gopal  
*University Of Connecticut*

Follow this and additional works at: <https://scholarsarchive.byu.edu/facpub>



Part of the [Management Information Systems Commons](#)

### Original Publication Citation

Singh, H., Kryscynski, D., Li, X. & Gopal, R. 2016. Pipes, Pools and Filters: How collaboration networks affect innovative performance. *Strategic Management Journal*. 37 (8): 1649-1666.

---

### BYU ScholarsArchive Citation

Singh, Harpeet; Kryscynski, David; Li, Xinxin; and Gopal, Ram, "Pipes, Pools and Filters: How Collaboration Networks Affect Innovative Performance" (2016). *Faculty Publications*. 1981.  
<https://scholarsarchive.byu.edu/facpub/1981>

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

## **Pipes, Pools and Filters:**

### **How collaboration networks affect innovative performance**

Harpreet Singh

Naveen Jindal School of Management  
University of Texas at Dallas

David Kryscynski

Marriott School of Management  
Brigham Young University

Xinxin Li

School of Business  
University of Connecticut

Ram Gopal

School of Business  
University of Connecticut

This is the peer reviewed version of the following article: Singh, H., Kryscynski, D., Li, X. & Gopal, R. 2016. Pipes, Pools and Filters: How collaboration networks affect innovative performance. *Strategic Management Journal*. 37 (8): 1649-1666, which has been published in final form at <http://onlinelibrary.wiley.com/doi/10.1002/smj.2419/full>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving."

## **Pipes, Pools and Filters:**

### **How collaboration networks affect innovative performance**

#### **ABSTRACT**

Research Summary: Innovation requires inventors to have both “new knowledge” and the ability to combine and configure knowledge (i.e. “combinatory knowledge”) and such knowledge may flow through networks. We argue that both combinatory knowledge and new knowledge are accessed through collaboration networks, but that inventors’ abilities to access such knowledge depends on its location in the network. Combinatory knowledge transfers *from* direct contacts, but not easily *from* indirect contacts. In contrast, new knowledge transfers from both direct and indirect contacts, but is far more likely to be new and useful when it comes from indirect contacts. Exploring knowledge flows in 69,476 patents and 89,930 unique inventors reveals evidence that combinatory knowledge from direct contacts and new knowledge from indirect contacts significantly affects innovative performance.

**KEYWORDS:** networks, innovation, tacit knowledge, pipes, knowledge pool, knowledge filter

Managerial Summary: Inventors often combine ideas to create innovations. To do this they need ideas to combine and they need the ability to combine those ideas. Inventors can get ideas to combine as well as the ability to combine ideas through prior co-workers. Prior co-workers can share ideas that may be relevant for the inventor’s project and can tell the inventor about other things that other people are working on, especially people the inventor may not know. This can help inventors easily learn about ideas from friends-of-friends. The ability to combine ideas, however, is much harder to pass on. Prior co-workers must carefully work with the inventor to teach her the complex processes of combining ideas. This means that it is very hard to learn how to combine knowledge from a friend-of-a-friend but it may be possible to learn from prior co-workers. We explore this phenomenon in the social relationships of software inventors.

## INTRODUCTION

Collaboration networks can enhance innovative performance by providing inventors with the two kinds of knowledge they need to innovate (Ahuja, 2000; Aral & Van Alstyne, 2011; Tortoriello *et al.*, 2014). First, network contacts can provide “new knowledge” that inventors can combine and reconfigure for innovations (Fleming, 2001; Mors, 2010; Rodan & Galunic, 2004). We define new knowledge as knowledge that is unfamiliar to the focal inventor and/or context and that is highly explicit. Second, network contacts may provide deep knowledge and abilities for how to combine and reconfigure new knowledge (Fleming, 2001; Henderson & Clark, 1990; Sutton & Hargadon, 1996). We define “combinatory knowledge” as the highly tacit ability and know-how to combine and reconfigure new knowledge. Both new knowledge and combinatory knowledge help inventors to create valuable innovations (Fleming, 2001; Hargadon & Sutton, 1997; Henderson & Clark, 1990; Nonaka, 1994; Nonaka *et al.*, 2000). Accordingly, networks are important for innovation because network contacts function as “pools” (Tortoriello *et al.*, 2014) of knowledge that can flow through the relational “pipes” (Podolny, 2001) of the social network.

While individuals function as knowledge pools in these networks, however, they may also function as “filters,” meaning that they may allow some knowledge types to pass through while preventing others (Ghosh & Rosenkopf, 2014). Specifically, knowledge and information that is highly explicit is relatively easy to pass from one person to another in a network (Hansen, 1999, 2002; Nonaka, 1994), and may not even require the willingness of the person who originally possessed the knowledge. In contrast, knowledge and information that is highly tacit may be difficult or even impossible to pass through individuals in a network (Szulanski, 1996). Passing on highly tacit knowledge requires deep investments by all parties (Hansen, 1999; Nonaka, 1994) and even if the person who originally holds the knowledge is willing to share that knowledge, the

transfer may still occur imperfectly (Collins & Guetzkow, 1964; Huber & Daft, 1987). Thus, individuals may act as filters preventing highly tacit knowledge from passing through.

The problem with viewing people as both pools and filters is that inventors may be limited in what knowledge they can gain from whom. Specifically, inventors may be able to acquire combinatory knowledge from direct contacts but not from indirect contacts. In contrast, however, they may be able to readily access new knowledge from both their direct and indirect contacts, but knowledge from indirect contacts is more likely to be new (Burt, 1992; Granovetter, 1973). Thus, the problem of explaining innovation in collaboration networks becomes far more complicated than simply explaining how knowledge flows in dyadic interactions between inventors (e.g. Mors, 2010; Rodan & Galunic, 2004). Rather, we need to understand what type of knowledge resides in the network, where that knowledge resides and how that knowledge flows, or not, as the case may be. Thus, the purpose of the present study is to articulate how combinatory and new knowledge available in different parts of the network affect innovative outcomes.

We argue, as we shall see, that the more valuable the combinatory knowledge available from direct contacts (i.e. the extent to which direct contacts have previously combined new knowledge successfully) the higher the focal inventor's innovative performance. We also argue, as we shall see, that new knowledge is more likely to come from indirect contacts. Our work joins the body of research exploring how networks affect innovation and contributes in at least two ways. First of all, we formalize the notion that individuals act as both pools and filters in knowledge networks. Accordingly we join a growing body of research emphasizing the potential frictions to knowledge flow in social networks (Ghosh & Rosenkopf, 2014; Schilling & Fang, 2014). Second, we combine the insights of Hansen (2002) suggesting that direct versus indirect paths in networks have different implications for knowledge flows with the work of Reagans and McEvily (2003) and others (Aral & Van Alstyne, 2011; Tortoriello *et al.*, 2014) suggesting that the kinds of knowledge available from

individuals in networks has important implications for innovative outcomes. Thus, we develop a more complete understanding of the pipes, pools and filters that affect innovation in collaboration networks.

### **COLLABORATION NETWORKS ENHANCE INNOVATIVE PERFORMANCE**

Innovation is the process of creating and putting into use new knowledge that results in value creation for stakeholders (Glynn, 1996). Innovation often entails taking existing solutions from one context and then combining them to create new solutions for the focal context (e.g. Henderson & Clark, 1990; Nonaka, 1994; Sutton & Hargadon, 1996). Inventors of these kinds of innovations thus need both: (1) “new knowledge” - i.e. solutions that are new and unfamiliar in the focal context and (2) “combinatory knowledge” – i.e. the knowledge and ability to combine solutions in new and useful ways. Inventors that have access to new knowledge but cannot combine that knowledge are unlikely to create anything useful. Similarly, inventors that have combinatory knowledge but lack new knowledge are unlikely to create anything novel. Accordingly, when inventors have both access to new and combinatory knowledge they are more likely to create valuable innovations. (Fleming, 2001; Hargadon & Sutton, 1997; Henderson & Clark, 1990; Nonaka, 1994; Nonaka *et al.*, 2000).

The dual importance of combinatory and new knowledge presents at least two problems for inventors seeking to create valuable innovations. First, combinatory knowledge may be very difficult to acquire. Rich descriptions of the innovative process suggest that combinatory knowledge has many tacit properties. Hargadon and Sutton (1997) for example, describe the deeply social process of inventors working together to share their implicit models for how to create new combinations. There is no codified handbook for how to combine solutions. While we certainly acknowledge that some aspects of combinatory knowledge may be explicit, our interpretation of these rich descriptions suggests that an inventor’s combinatory knowledge is largely tacit – i.e. very

sticky and difficult to transfer (Hansen, 1999). Accordingly, inventors gain combinatory knowledge through personal experimentation, learning-by-doing and/or through complex social interactions and socialization processes (Nonaka, 1994).

One important way inventors can gain combinatory knowledge is by learning from those who have been successful at innovating before. These successful inventors likely have complex mental models, frameworks and ways-of-doing that they can pass on to the focal inventor through intense interpersonal interactions. The problem, however, is that these successful inventors are not likely to engage in the intense communication needed to transfer such tacit knowledge with people they do not know well (Daft & Lengel, 1986). This means that focal inventors may only be able to learn combinatory knowledge from successful inventors with whom they have strong interpersonal relationships. Without such strong relationships, their only way to improve their combinatory knowledge may be through their own experience.

Second, there is too much new knowledge in the world for inventors to efficiently scan and access in their innovative efforts. Rich descriptions of the innovative process suggests that much of the new knowledge accessed by inventors has explicit properties. Again, Hargadon and Sutton (1997) describe how inventors are able to see solutions in one space and gain explicit knowledge about those solutions simply from seeing them. An inventor, for example, could see a software solution and infer the underlying architecture without gaining the deep tacit know-how the inventors of that software needed in order to create it in the first place. Inventors can gain access to products, articles, patents and so forth that contain codified solutions that they can use in the innovative process. While we acknowledge that some of the solutions required for innovations may have tacit properties, our interpretation of these rich descriptions suggests that much of the new knowledge acquired for innovation is highly explicit. This new knowledge is often easy to acquire through reading a paper, taking apart an electronic device, or observing a product's architecture, but there is

too much new knowledge in the world for inventors to efficiently explore. Rationally bounded inventors cannot search all possible new knowledge in their efforts to find what is relevant (Fleming, 2001; March & Simon, 1958).

Interpersonal collaboration networks, or connections between individuals who have co-invented together in the past, may help inventors with both of these problems because prior collaborators represent “pools” of knowledge resources that can be accessible through the relational “pipes” of the network. These prior collaborators may have both combinatory knowledge that they can share with the focal inventor as well as new knowledge that they can pass on to the inventor. Given the generally intense interactions between co-inventors (Fleming *et al.*, 2007; Singh, 2005) it is likely that collaborators have both the motivation and the shared experiences to transfer highly tacit combinatory knowledge to the focal inventor. Similarly, prior collaborators may provide an efficient means of searching the broad and expansive knowledge landscape to find relevant new knowledge (Borgatti & Cross, 2003; Cohen & Levinthal, 1990). Rather than searching all available databases, inventors can discuss their projects with prior collaborators who may know of related information in the broader knowledge landscape (Fleming *et al.*, 2007). Thus, interpersonal collaboration networks may provide both a means for focal inventors to enhance their combinatory knowledge as well as efficient access to new knowledge.

While it seems clear that collaboration networks may help, prior work has either examined knowledge heterogeneity without exploring combinatory knowledge (e.g. Hansen, 2002), explored both knowledge heterogeneity and combinatory knowledge but without exploring network relationships beyond direct contacts (e.g. Mors, 2010), or examined how knowledge flows through different types of ties but without examining innovative performance as the key outcome of interest (e.g. Hansen, 2002). Thus, while prior work has clearly laid the foundation for our arguments, we do not yet have

insights regarding the optimal network configurations for inventors who need access to both diverse new knowledge and combinatory knowledge for their innovative performance.

## **UNDERSTANDING DIFFERENT KNOWLEDGE TYPES IN DIFFERENT LOCATIONS**

We combine prior insights on how the knowledge residing in direct contacts affects innovative performance (Mors, 2010; Rodan & Galunic, 2004) with Hansen's (2002) insights on how direct and indirect paths differ for knowledge transfer to argue that different types of knowledge matter more for innovative performance when they reside in different network locations.

### **Combinatory Knowledge from Direct Contacts**

Combinatory knowledge should have a greater impact on an inventor's innovative performance when it resides in the inventor's direct rather than indirect contacts. Direct paths in a collaboration network represent relationships between two inventors who have previously collaborated on a project. Inventors who lack combinatory knowledge may be able to gain it through their interactions with prior collaborators who have it (Hansen, 2002; McFadyen & Cannella, 2004; Sorenson *et al.*, 2006). Inventors can reach out to these collaborators to ask for help and insight. These collaborators can pass on valuable knowledge, insights, frameworks and experiences about how to combine solutions and the processes they have used in the past (Tortoriello *et al.*, 2014) – i.e. they can pass on valuable combinatory knowledge to the focal inventor. We therefore expect that an inventor's innovative performance on a given project will be positively associated with the value of combinatory knowledge that exists among her direct contacts – i.e. the extent to which her direct contacts have successfully combined knowledge in prior projects to achieve high levels of innovative performance. Thus:

*Hypothesis 1a: The more valuable the combinatory knowledge among an inventor's direct contacts in a collaboration network, the higher the inventor's innovative performance.*

While combinatory knowledge from direct contacts may have a significant positive impact on the inventor's innovative performance, these direct contacts are not likely to provide much new knowledge for the inventor to combine and reconfigure. The knowledge landscape of direct contacts is rarely new to the focal inventor because as people work together and have shared experiences, they tend to naturally develop an understanding of what others know (Aral & Van Alstyne, 2011). When collaborators work together intensely over extended periods of time, they likely develop detailed understandings of each other's knowledge profiles. While they do not likely internalize all that their collaborators know and understand, they do gain a deep understanding of the breadth and depth of collaborators' different kinds of knowledge (Liang, Moreland, & Argote, 1995). We likely experience this in co-author relationships in academia. When we work closely with our co-authors we reveal to each other our perspectives, paradigms, training, and experiences simply through our conversations. A network theorist collaborating with an information systems (IS) researcher, for example, will not gain much proficiency in the profession specific knowledge required for information systems, but will learn a great deal about what that profession specific knowledge is. In other words, she gains a rich understanding of the knowledge profile of her IS colleague without personally gaining and incorporating that knowledge profile. In other words, her IS colleague's knowledge is no longer "new" to her.

While it is quite difficult to precisely identify what knowledge is new and what knowledge is previously known by an inventor, scholars have used the heterogeneity of knowledge in a social network to proxy for the availability of new knowledge in that network (Rodan & Galunic, 2004). Prior arguments suggest that heterogeneity indicates the probability that some knowledge in the network will be truly new to the inventor. Thus, as knowledge heterogeneity in the network increases, the innovative performance of the inventor should also increase (Rodan & Galunic, 2004). While this may be true generally, our arguments suggest that the probability of finding new

knowledge is likely very low among direct contacts regardless of how heterogeneous their knowledge may be. Even very heterogeneous knowledge among these direct contacts is likely part of the focal inventor's knowledge landscape and, therefore, not new to the focal inventor.

Thus, while combinatory knowledge of direct contacts may have a significant positive impact on an inventor's innovative performance per the logic preceding hypothesis 1a, knowledge heterogeneity among those direct contacts may not have much of a positive impact. Thus:

*Hypothesis 1b: The value of combinatory knowledge available from direct contacts has a stronger effect on an inventor's innovative performance than the heterogeneity of knowledge available from those direct contacts.*

### **New Knowledge from Indirect Contacts**

While direct paths seem most useful for providing valuable combinatory knowledge, indirect paths are likely most useful for providing useful new knowledge. As mentioned previously, the inventor generally knows her direct contacts' knowledge profiles, but may not know her indirect contacts' knowledge profiles. Her direct contacts, however, know the knowledge portfolios of their direct contacts, a.k.a the focal inventor's indirect contacts. Thus, the inventor's direct contacts are well positioned to pass on information about what their direct contacts know when the focal inventor could benefit from that knowledge (Lin, 2001; Sorenson *et al.*, 2006). Prior work supports the notion that these kinds of indirect contacts can provide timely access to needed knowledge (Burt, 1992; Hansen, 1999, 2002; Sorenson *et al.*, 2006).

The new knowledge shared by contacts is often embodied in documents, products, artefacts and so forth (Hargadon & Sutton, 1997). This means that transferring the new knowledge could be as simple as showing the inventor an article, website, product, and so forth. Consider how this may work for co-author networks in management research. A researcher may be struggling with a certain empirical problem and reach out to direct contacts (prior co-authors) for help. None of the prior co-

authors know how to solve the empirical problem, but one of the co-authors has a prior co-author (an indirect contact to the focal researcher) who has a paper that solves a similar empirical problem. Thus, the direct contact may share her prior co-author's paper with the focal researcher. She does not need to deeply understand it, nor does she need to inform her prior co-author that she is sharing it. She simply needs to provide a reference to this published work so that the focal researcher can read it and learn from it.

The key point of this discussion is that collaborators can quickly and easily transfer new knowledge due to their relationships with their direct contacts. They are not likely to pass on all of the available new knowledge, but when an issue comes up in conversation they can provide referrals to relevant knowledge. This means that the focal inventor need not personally search all available knowledge in the external environment, but can leverage her direct contacts' direct contacts for an efficient search for relevant new knowledge.

We argued above that heterogeneous knowledge among direct contacts is likely redundant for the focal inventor and, therefore, not particularly useful for the inventor's innovative performance. In contrast, however, knowledge heterogeneity among indirect contacts is probably not redundant. It is likely that heterogeneous knowledge among indirect contacts does provide an indicator of the unfamiliar solutions available to the focal inventor through those contacts. Accordingly, heterogeneous knowledge among indirect contacts likely enhances the innovative performance of the focal inventor by enhancing the inventor's efficient access to new and unfamiliar solutions. Thus:

*Hypothesis 2a: The more heterogeneous the knowledge available from a focal inventor's indirect network contacts, the higher the inventor's innovative performance.*

An inventor's indirect contacts may also have valuable combinatory knowledge, but this knowledge may not be accessible to the focal inventor because of its tacitness. Combinatory

knowledge may not flow effectively from indirect contacts through direct contacts to the focal inventor. In other words, direct contacts may function as “filters” in the pipes connecting a focal inventor with an indirect contact due to potential frictions in knowledge flows in networks (Ghosh & Rosenkopf, 2014). Specifically, the direct contact facilitates the easy transfer of important new knowledge, but blocks the passage of highly tacit combinatory knowledge.

There are several reasons why direct contacts may act as filters that block the passage of combinatory knowledge. First, the direct contacts in this case act as knowledge intermediaries – i.e. they simply pass knowledge on from one person to another. As discussed previously, codified new knowledge can be transferred easily. Thus, passing along new knowledge is a very low cost way for a direct contact to help a prior collaborator. Combinatory knowledge, in contrast, must be learned through doing and/or deep socialization. Learning combinatory knowledge requires a significant investment both on the part of the transferor and the transferee. This means that to pass on such knowledge the direct contact must be willing to (1) make a deep personal investment in gaining the knowledge of one of her direct contacts and then (2) be willing to then deeply invest in passing that knowledge along to the focal inventor. Additionally, the indirect contact must be willing to invest in transferring the knowledge to the direct contact. The indirect contact may be willing to pass on combinatory knowledge for the benefit of the direct contact, but may be less motivated to make such an investment just so the direct contact can pass it on to someone else.

Second, even if all parties are willing to make investments in transferring combinatory knowledge, there still may be information loss in transmission. Intermediaries might forget details and fail to mention all that they have learned from indirect contacts of the focal inventor (Collins & Guetzkow, 1964; Huber & Daft, 1987). Lacking direct interaction, the focal inventor will not have the opportunity to seek comments, feedback, or critique, and hence will not be able to completely assimilate the tacit knowledge of indirect contacts. Thus, individuals may act as filters for

combinatory knowledge even if they are fully willing to pass such knowledge along.

In summary, then, direct contacts in a focal inventor's networks likely function as filters in the knowledge network that allow new knowledge to pass through easily, but that block the flow of combinatory knowledge. This filter effect means that indirect contacts are much more valuable to the focal inventor due to their knowledge heterogeneity rather than the value of their combinatory knowledge. While these indirect contacts may possess valuable combinatory knowledge, it simply does not flow as well through the direct contact in the network. Thus:

*Hypothesis 2b: The heterogeneity of knowledge available from indirect contacts has a stronger effect on an inventor's innovative performance than the value of combinatory knowledge available through those indirect contacts.*

## METHODS

We test our hypotheses using patent data from the knowledge intensive software industry. Software patents, as an indicator of software innovation, protect the intellectual property critical to the returns on inventors' investments. The creation of software innovations is dependent on a large number of software technologies and typically requires groups of inventors with diverse forms of knowledge to collaborate. Software patents do not fully disclose the relevant code when they are issued (Hall & MacGarvie, 2009), so connections with inventors in the collaboration network who have worked on patents related to the new innovation can be valuable. This makes software an interesting context in which to study the relationship between interpersonal collaboration networks and innovative performance.

Our data were collected from the United States Patent and Trademark Office (USPTO) homepage using the approach developed by Hall and MacGarvie (2009). Our initial sample contains patents for which applications were submitted between 1976 and 1997 and that were granted by 2004. This gave us 69,476 patents and 89,930 unique inventors. In our dataset, each patent record

contains the patent number, application date (the date on which the patent was filed at the USPTO), grant date (the date on which the patent was granted), the names of the inventors, their addresses, technology classes (we use the three-digit primary technology classification defined by the USPTO to indicate the technology class of the patent), owner firm and claims. For each patent, our dataset also contains the list of all of the patents that it cites (backward citations) and all of the later patents that cite it (forward citations) until November 2007 (the time when the data were collected). We supplemented our patent data with unique inventor identifiers provided by Lai et al. (2009),<sup>1</sup> firm identifier data from the NBER database and firm R&D expenditure and sales data from the Compustat database.

Our sample is likely to contain some firms that are unusually good at creating highly innovative patents for reasons that we cannot observe in our data. For this reason we needed to carefully match patents to firms. This is problematic because patents do not always list the firm name consistently. For example, a patent assigned to Microsoft may list the firm as “Microsoft”, “Microsoft Corporation”, or “Microsoft India”, or it may list a name that is not associated with Microsoft at all. To address this issue, we merged the patent data from the USPTO with the dataset from Hall et al. (2001) and used assignee codes in their dataset to map alternative assignee names to a single firm identifier. Additionally, some patents are listed under the name of a subsidiary rather than the parent firm. To address this we used Compustat-based firm identifiers from the NBER database to match approximately 1,400 assignees to their parent firms and used the Lexus Nexus database, the Goliath database and the other internet databases to match another 900 assignees to their parent firms. These approaches together allowed us to match approximately 2,300 assignees to 1,425 unique organizations. These 1,425 unique organizations owned 52,832 patents (or about 76 percent of all

---

<sup>1</sup> We have also used the algorithm developed by Singh (2005) to identify unique inventors and to construct our network. The results from this robustness check are qualitatively similar to those reported in our paper.

assigned patents). It was not feasible to identify owners for the remaining 24 percent of the patents, which belonged to around 6,700 assignees, each of whom owns only a few patents. Because our models used firm fixed effects to control for unobserved firm-level heterogeneity, we required sufficient observations for each firm and therefore restricted our sample to patent applications from assignees with at least ten patents. In this way, we obtained a final dataset of 50,066 patents, 459 firms and 60,974 inventors.<sup>2</sup>

Since our data comes from patents, the patent becomes the unit of analysis in our empirical strategy. We ultimately predict the performance of a particular patent using variables that proxy for what collaborators know and how those collaborators are connected. As we shall see, what an inventor knows is measured by characteristics of prior patents s/he has co-invented and how the inventor is connected is measured by prior co-inventor relationships. These measures are discussed in detail below.

### **Dependent Variable**

Our theory predicts the innovative performance of inventors based on their collaboration networks, but patent data do not allow us to identify the innovative performance of individual inventors unless they are solo inventors on a particular patent project – a rarity in our sample. We thus use the patent as the unit of analysis and focus on the innovative performance of the patent as a proxy for the innovative performance of the co-inventors on that project. Thus, the key dependent variable in our study is the innovative performance of a patent, measured as the forward citations of the focal patent excluding self-citations<sup>3</sup>. Prior work suggests that forward citation counts are an appropriate indicator of the significance of innovations and can thus be used to measure the innovative performance of patents (Albert *et al.*, 1991; Fleming, 2001; Hall *et al.*, 2005).

---

<sup>2</sup> We also performed the analysis using cut-off points of 5 and 20 with no substantive change to our main findings.

<sup>3</sup> We include self-citations in a robustness check and find no substantive difference in our results.

## Independent Variables

**Direct and Indirect Contacts.** Direct contacts are those in an inventor's collaboration network with a path distance of 1 and indirect contacts are those with a path distance of 2. We used a "whole-network" approach to construct the collaboration network, which relies on the criterion of membership to define the network boundaries. Using this approach, we included all of the inventors of software patents in our network, i.e., we used participation in proprietary software innovation as the network boundary. This is consistent with previous studies using membership in an industry (Schilling & Phelps, 2007), a region (Fleming *et al.*, 2007) or a technology platform (Singh *et al.*, 2008) to construct networks.

A collaboration tie forms when two inventors work together on a project, and this is operationalized in our data as the first time two inventors appear together on a patent application (Fleming *et al.*, 2007; Nerkar & Paruchuri, 2005; Singh, 2005, 2008). Consistent with prior work, we assume that the collaboration begins with the first project and that the interpersonal relationship persists long after the completion of that project (Fleming, 2007). We followed Nerkar and Paruchuri (2005) in creating collaboration networks for the inventors using three-year moving windows. Similar approaches have been used in the literature to analyze other collaboration data, with windows ranging from one to five years (Fleming *et al.*, 2007; McFadyen & Cannella, 2004; Schilling & Phelps, 2007). This approach uses patents within each three-year window to construct the collaboration network, and this collaboration network is used to predict the performance of patents applied for in the following year. For example, we examined how network ties evident during the 1976-1978 period affected the value of the patents applied for in 1979, how network ties evident during the 1977-1979 period affected the value of the patents applied for in 1980, and so on.

Figure 1 illustrates the construction of an interpersonal inventor collaboration network. Based on the approach suggested by Wasserman and Faust (1994), we used a fully linked clique to

represent collaborators on a particular patent; the fully linked cliques are connected to each other when an inventor participates in the development of multiple patents. For instance, in Figure 1, the top row depicts four patents and the inventors who worked on each patent; for instance, A, B and C worked on the first patent, whereas C and D worked on the second patent. Co-inventors of the same patent are members of a fully linked clique, and thus, there are four cliques in the second row (ABC, CD, DEF and GH) for the four patents. Because C and D worked on multiple patents, the first three cliques are further connected, as shown in the third row. As new patents are started, more cliques appear and join the network. The bottom row shows an example of how these networks evolve over time. Each affiliation network is used to create a unipartite projection for the inventor network, as shown in Figure 1. Following Borgatti et al. (2002), we used UCINET 6 to obtain measures for these networks.

[INSERT FIGURE 1 ABOUT HERE]

**Value of Combinatory Knowledge.** Combinatory knowledge is the knowledge and ability to combine and reconfigure knowledge from different technical areas into valuable innovations. As argued previously, this combinatory knowledge is likely either gained by doing and/or gained through intense socialization processes (Nonaka, 1994). So, an inventor who works on a project that integrates knowledge from many technical areas likely gains this experience of combining different knowledge elements.

Of course, not all knowledge combinations result in valuable innovations. Some inventors may combine solutions in ways that do not create significant value for any stakeholders. These inventors may know how to combine solutions, but may not know how to do so in useful and valuable ways. In contrast, inventors who have developed successful innovations likely develop the ability to create value for stakeholders. However, inventors might also have created successful innovations that do not require any combinatory knowledge. Therefore inventors on patents that combine knowledge

across technical areas as well and that are highly cited are more likely to possess valuable combinatory knowledge.

We are not aware of prior work measuring combinatory knowledge through patents so we construct a novel measure of combinatory knowledge as follows. We use the proportion of patents in an inventor's personal portfolio that are both highly valuable and high in combinations as our measure of valuable combinatory knowledge. Each patent is assigned to many technical classes and the number of technical classes indicates the number of different technical domains that were combined in its creation. The value of the knowledge embedded in the patent is measured through normalized forward citations<sup>4</sup>. We first categorize each patent as being *hi-co* if both the forward citations received and number of combinations are more than the fiftieth percentile (the results are robust to multiple percentile cutoffs). The value of a contact's combinatory knowledge is then measured as the proportion of *hi-co* patents in that contact's portfolio of patents filed during the three-year moving window prior to the focal innovation being created. The average combinatory knowledge of all the direct (indirect) contacts then gives the value of direct (indirect) contacts' combinatory knowledge available to all co-inventors on the focal patent.

**Knowledge Heterogeneity.** Knowledge heterogeneity is measured as the diversity of *unfamiliar technical classes* in the prior patents of network contacts. We define *unfamiliar technical classes* as technical classes on which direct and indirect contacts have worked but co-inventors of focal patents have not worked in the past. In other words, we use technical classes which are *unfamiliar*

---

<sup>4</sup> Forward citations suffer from a truncation bias because we observe forward citations for a shorter time period for patent applications submitted later than for those submitted earlier, which may bias our measure. To address this issue, we constructed our measure with normalized forward citation figures (i.e., forward citation counts divided by the average forward citation count for all the patents issued during the same application year and in the same primary three-digit technical class).

to co-inventors of focal patent to calculate knowledge heterogeneity<sup>5</sup>. To measure the knowledge heterogeneity of direct contacts, we use the distribution of technical classes of patents on which the direct contacts have worked in the three-year moving window prior to the focal innovation being created. The measure of *direct contacts' knowledge heterogeneity* is calculated as one minus the Herfindahl of concentration of technical classes of direct contacts' patents- a well-established measure of diversity (Avenel *et al.*, 2007; Grant *et al.*, 1988; Tallman & Li, 1996). Specifically this is calculated as:  $1 - \sum_t \left(\frac{n_t}{n}\right)^2$ , where  $n_t$  is the number of unique patents applied for by direct contacts in the last three years prior to the focal innovation being created which belong to the unfamiliar technical subclass  $t$ ,  $n$  is the total number of patents applied by the team's direct contacts in all the unfamiliar technical classes in the last three years. The measure of *indirect contact's knowledge heterogeneity* is identical, but based on the patents applied for by indirect rather than direct contacts.

### **Control Variables**

We controlled for numerous attributes of the co-inventors, the focal patents, and the region to address alternative explanations and potentially confounding factors. These control variables are defined and described in Table 1 below:

[INSERT TABLE 1 ABOUT HERE]

### **Model Specification**

The dependent variable in this model, the number of forward citations, is a count variable that takes only non-negative integer values and has a skewed distribution. Utilizing a linear regression

---

<sup>5</sup> Our thanks to an anonymous reviewer who pointed out that our measure assumes that the knowledge distance between technical classes is similar for all technical classes. This is a limitation of our measure as we did not have the data to explore the actual knowledge distance between technical classes or, put differently, the extent of overlap between classes. Nevertheless we do believe our measure proxies for the probability of finding new knowledge in the network.

can result in inconsistent, biased and inefficient estimates (Greene, 2003). A Poisson regression is a common approach to model quantitative data (Hausman & Bronwyn, 1984). However, such a model makes a strong assumption of equal mean and variance. If over-dispersion is present, the standard errors of coefficients will be underestimated, leading to spuriously high levels of significance (Cameron & Trivedi, 1986). Because our data demonstrate over-dispersion (rejecting the Poisson model at  $p < 0.0001$ ), we utilized a negative binomial regression for our analysis. This method is a generalization of the Poisson model and allows for over-dispersion.

Table 2 provides the descriptive statistics for all of the variables. We performed logarithmic transformations on the variables that are right skewed<sup>6</sup> and square transformations on the variables that are left skewed (Gelman & Hill, 2007). The variable transformations are also provided in Table 2. A correlation matrix for the transformed variables is shown in Table 3. Multicollinearity does not appear to be an issue. As a further diagnostic test, we calculated the variance inflation factors (VIF) for each variable (Greene, 2003); the maximum VIF is below the generally accepted threshold of 10 (Belsley *et al.*, 2004). We also addressed the following specification issues in our empirical analysis. First, we used the Huber White robust standard errors to correct for potential heteroskedasticity (Greene, 2003). Second, to alleviate the concern of reverse causality (i.e., the idea that the team's performance shapes the network, not vice versa), we constructed our network measures one year prior to the focal patent's application year because patents start to receive citations (the dependent variable) after their application date. Unobserved or immeasurable attributes of observationally equivalent firms may potentially influence the quality of innovations. Some firms may simply be better able to create highly innovative patents. Our model accounts for these firm-specific unobserved effects with firm fixed effects. Patents applied for in different years

---

<sup>6</sup> If some observations have zero values for a variable, we added one before making logarithmic transformations.

will have different observed time spans during which they will receive citations. To account for systematic cross-year differences in citation patterns, we included application year fixed effects in our regressions. Patents belonging to different technical categories differ in the number of citations received (Hall *et al.*, 2001). To control for the systematic difference in citation patterns across *technological classes*, we also included fixed effects for three-digit technology primary classes of patents.

[INSERT TABLES 2 AND 3 ABOUT HERE]

## Results

Table 4 illustrates the difference in means in our dependent variable<sup>7</sup> based on splits in our key independent variables. For each variable the “high group” is those one standard deviation above the mean and the “non-high” group comprises the remaining observations. The patterns in table 4 are consistent with our hypotheses.

The results of the negative binomial regressions are presented in Table 5. Model 1 presents the base model with only the control variables included. Model 2 adds to Model 1 the knowledge heterogeneity of both direct and indirect contacts. Model 3 adds to Model 1 the value of combinatory knowledge from both the direct and indirect contacts. As predicted, the effect of direct contacts’ combinatory knowledge is positive and significant lending support to hypothesis 1a. As predicted, the effect of knowledge heterogeneity among indirect contacts is positive and significant lending support to hypothesis 2a<sup>8</sup>. We performed likelihood ratio tests to examine whether models 2 and 3 significantly improved the explanatory power of the model over the controls only model and

---

<sup>7</sup> Forward citations differ significantly across technical classes. To address this issue, we constructed our measure with normalized forward citation figures (i.e., forward citation counts divided by the average forward citation count for all the patents issued during the same application year and in the same primary three-digit technical class).

<sup>8</sup> An anonymous reviewer astutely pointed out that prior work suggests a non-linear effect of knowledge heterogeneity. We suspect such non-linearity may be due to the conceptual difference between direct and indirect ties, but we have also tested for non-linear effects. We do not see any evidence of non-linear effects in our results.

the resulting chi-squares were statistically significant – i.e. the key independent variables significantly increased the explanatory power of the model.

Model 4 includes both knowledge heterogeneity and the value of combinatory knowledge for direct and indirect contacts. In Model 4, consistent with the findings above, the value of direct contacts' combinatory knowledge and the heterogeneity of indirect contacts' knowledge are both positive and significant. However to test H1b, we need to test the null hypothesis H0:

$\beta_{combinatory\_direct} - \beta_{heterogeneity\_direct} \leq 0$  against the alternative hypothesis

Ha:  $\beta_{combinatory\_direct} - \beta_{heterogeneity\_direct} > 0$ . A Wald test rejects the null hypothesis ( $p$ -value  $< 0.001$ ) which indicates that effect of direct contacts' combinatory knowledge is significantly greater than effect of knowledge heterogeneity of direct contacts. This test lends support to H1b.

Similarly for H2b, we need to test the null hypothesis H0:  $\beta_{heterogeneity\_indirect} -$

$\beta_{combinatory\_indirect} \leq 0$  against the alternative hypothesis Ha:  $\beta_{heterogeneity\_indirect} -$

$\beta_{combinatory\_indirect} > 0$ . A Wald test rejects the null hypothesis ( $p$ -value  $< 0.001$ ) which indicates that effect of knowledge heterogeneity among indirect contacts is significantly greater than effect of indirect contacts' combinatory knowledge. Thus, hypotheses 1b and 2b are both supported.

To assess the economic significance of our results, we also compute the magnitude of these effects using Hall *et al.* (2005) as a source of reference. Hall *et al.* (2005) find that a firm's market value increases by 3% for a unit increase in the average number of citations to its patents. They further point out that it is not easy to increase the average forward citation counts by one. Our finding thus has important implications for firms looking to enhance their market value by improving their employees' social capital. Take a median firm in our dataset whose patents get on average 23 citations per patent. For a negative binomial model,  $E(Y|X) = \exp(X\beta)$ , where  $E(Y|X)$  is the expected value of dependent variable  $Y$ , given independent variables,  $X$ . If independent variables are log transformed then  $a\%$  increase in independent variable corresponds to  $100 * a\%$

$\left(\left(1 + \frac{a}{100}\right)^\beta - 1\right)\%$  increase in Y. For combinatory knowledge of direct contacts,  $\beta = 0.13$ , therefore a 10% increase in average value of combinatory knowledge corresponds to 1.24 % increase in the number of forward citations. This corresponds to an equivalent increase of 0.30 citations for this firm which correspond to a 0.86% increase in the firm's market value. Similarly, a 10% increase in average knowledge heterogeneity among indirect contacts corresponds to 0.8% increase in the firm's market value. While these estimated effects on firm value are based on aggregate estimates and should be treated with caution, they highlight the potential economic significance of our findings.

[INSERT TABLES 4 and 5 ABOUT HERE]

### **Robustness Checks**

**Alternative Measures.** While we believe our measures are reasonable proxies for an inventor's access to combinatory and new knowledge, we also tested a number of alternative constructions as potential proxies. For combinatory knowledge we tested: (1) the number of a contact's *hi-co* patents (2) the product of the average number of combinations and the average forward citation counts of patents, and (3) the average number of forward citations and average number of combinations respectively for contacts' patents in last three years. For knowledge heterogeneity we tested: (1) technical classes in which contacts have worked in three years prior to focal innovation being created to calculate knowledge heterogeneity rather than only using *unfamiliar technical classes*, (2) the number of unique technical classes on which contacts have worked in past three years but on which co-inventors of focal patents have not worked in the past, and (3) the number of unique technical classes on which contacts have worked in past three years. The results for all of these alternative measures were substantively consistent with the main findings.

**Endogeneity.** Because network variables and innovative performance may both be driven by unobserved co-inventor quality, endogeneity is a potential concern in this type of study. To address

this issue, we first used co-inventors' past performance to control for co-inventor's ability (Schilling & Phelps, 2007). In addition, we took the hedonic approach suggested by (Reagans *et al.*, 2007), to establish exogeneity for collaborations in this context. The results were qualitatively similar.

**Star effects.** It could be argued that co-inventors that have worked with well cited inventors in the past might receive more citations because of the affiliation to the “star inventor”. To address this concern we did an additional robustness check to account for the presence of ‘star inventors’ in the focal inventors’ network. The main results are qualitatively similar even after controlling for star effect. This implies that our model has explanatory power above and beyond any star effects.

**Interaction effects.** The logic embedded in our manuscript suggests that both new knowledge and combinatory knowledge are important for innovation. We have hypothesized the main effects of each, and we have hypothesized that heterogeneous knowledge is more valuable when it comes from indirect contacts while combinatory knowledge is more valuable when it comes from direct contacts. This logic implies that there may be an interaction between combinatory knowledge that comes from direct contacts and heterogeneous knowledge available from indirect contacts. We tested this possibility but the coefficients on the interaction term were not significant in any of our specifications. While our data do not allow us to carefully explore the potential explanations for why we do not find evidence for an interaction effect, it is possible that the primary source of combinatory knowledge is the collaborations with other inventors, but inventors may have access to diverse and new knowledge through many other sources outside of the inventor collaboration network. If true, then we may not be able to measure the true access to new knowledge but, instead, only measure access to diverse knowledge available within this one network type. A more expansive network study that could pick up new knowledge from other sources as well may be able to pick up such an interaction effect.

**Other Robustness checks.** We performed additional tests for multi-collinearity by randomly omitting 1/3 of the observations in the analyses and found no substantive change (Greene, 2003). We also tested: (1) alternative time windows of two and four years for network calculations, (2) adding time and technology interaction effects, (3) using clustered robust standard errors (based on a firm identifier), and (4) analysis without excluding self-citations from the calculation of forward citations. The results from these analyses are qualitatively similar to the results of the main model.

## **DISCUSSION AND CONCLUSION**

We have explored the importance of valuable combinatory knowledge among an inventor's direct contacts and the importance of heterogeneous knowledge among an inventor's indirect contacts for innovative performance. Our results suggest that both the value of combinatory knowledge and the heterogeneity of knowledge available through network contacts have significant influence on the performance of co-inventor projects and that their influences vary for contacts at different distances. Specifically, the value of direct ties' combinatory knowledge has a much stronger impact on innovative performance than the heterogeneity of those direct contacts' knowledge. Similarly, the heterogeneity of indirect ties' knowledge has a much stronger impact on innovative performance than the value of their combinatory knowledge. These results provide support for the idea that individuals in collaboration networks may function as filters in the knowledge network allowing some types of information to pass freely through while preventing other types of knowledge from transferring.

The theoretical implications of our study suggest the need to explore simultaneously the resources held by actors in a social network as well as how those actors are connected to the focal unit. Some resources may freely flow through individuals while others may not. We thus join a growing body of research exploring the potential frictions in knowledge flows through social networks (Ghosh & Rosenkopf, 2014; Schilling & Fang, 2014). Our work suggests that

combinatory knowledge may not effectively pass through individual nodes in social networks, but that new knowledge may flow very quickly and efficiently. Thus, studying network structure cannot sufficiently explain performance outcomes without carefully tracking what knowledge can flow from different parts of the structure. Similarly, studying the knowledge available in the network cannot sufficiently explain performance outcomes without carefully mapping out the pathway connecting that knowledge to the focal unit. Thus, our work further emphasizes the need to study the resource pools and the connecting pipes of social networks simultaneously.

One result that seems particularly important to mention is that our results do not support the Rodan and Galunic (2004) hypothesis and findings that heterogeneous knowledge among direct contacts enhances innovative performance. There are several potential explanations for this inconsistency that are worth exploring in future work. First of all, Rodan and Galunic (2004) focus on managers in a telecommunications company rather than inventors. Accordingly, the nature of innovation for operational managers may be different from the nature of innovation for our inventors. Second, these authors use managers in the telecommunications company who likely have wide variance in their innovative performance. Our inventors are all successful enough that they have patented at least once in the data, and many of them multiple times. It is possible that Rodan and Galunic's wider variance in the innovative performance of managers highlights the fine grained importance of knowledge heterogeneity even among direct contacts that our coarser measures could not pick up. Third, it is also possible that Rodan and Galunic's survey measure to construct knowledge heterogeneity actually picks up the indirect effect of indirect contacts' heterogeneous knowledge. The focal manager indicates how different/similar each contact is to other direct contacts. When making this assessment, the manager may not be able to distinguish clearly between knowledge that resides in the direct contact versus knowledge that comes through the direct contact from indirect contacts. Accordingly, it is possible that the significant effects of

knowledge heterogeneity in their study are actually picking up the true knowledge heterogeneity among their indirect contacts. If so, then our results are more consistent than they appear at first glance.

While our results suggest interesting implications, the current analysis has several limitations. First, since we use historical patent data, we cannot directly observe social interactions (Nerkar & Paruchuri, 2005; Sorenson *et al.*, 2006). Additionally, we theorize that combinatory knowledge and new knowledge flow differently through collaboration networks, but we cannot directly observe either combinatory or heterogeneous explicit knowledge. Future research may benefit from more direct observations of relationships and social interactions, as well as more direct measures of combinatory and heterogeneous knowledge in an innovation context.

Second, our results are based on a single type of interpersonal relationship, which may capture only a fraction of interpersonal contacts. Many of the inventors we study may know each other through other professional means. Similarly, many of our inventors have social relationships with others who are not patenting inventors who likely have strong influences on their innovative outcomes. Thus, future research may benefit from exploring other types of relationships in addition to patent collaborations, yielding a more comprehensive understanding of how social networks affect innovations.

Third, we drew our conclusions by studying software patents, and software may be quite distinct from other contexts such as biotechnology. Fourth, we used forward citations to measure the value of software innovations. Further studies may examine whether the results obtained in this study still hold if other measures, such as patent renewals and the number of countries in which a patent is filed, are used to measure the value of innovations.

## References

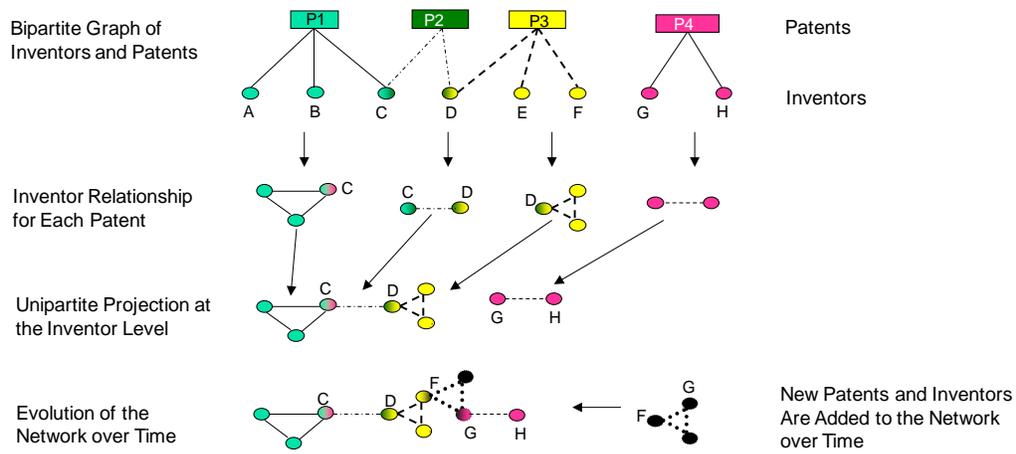
- Ahuja G. 2000. Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly* **45**(3): 425-455.
- Albert MB, Avery D, Narin F, McAllister P. 1991. Direct validation of citation counts as indicators of industrially important patents. *Research Policy* **20**(3): 251-259.
- Aral S, Van Alstyne M. 2011. The diversity-bandwidth trade-off1. *American Journal of Sociology* **117**(1): 90-171.
- Argyres NS, Silverman BS. 2004. R&D, organization structure, and the development of corporate technological knowledge. *Strategic Management Journal* **25**(8-9): 929-958.
- Avenel E, Favier A-V, Ma S, Mangematin V, Rieu C. 2007. Diversification and hybridization in firm knowledge bases in nanotechnologies. *Research Policy* **36**(6): 864-870.
- Belsley DA, Kuh E, Welsch RE. 2004. *Regression diagnostics: Identifying influential data and sources of collinearity*. Wiley-IEEE: New York.
- Borgatti SP, Cross R. 2003. A relational view of information seeking and learning in social networks. *Management science* **49**(4): 432-445.
- Borgatti SP, Everett MG, Freeman LC. 2002. *Ucinet for Windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.
- Burt RS. 1992. *Structural holes: The social structure of competition*. Harvard University Press: Cambridge, MA.
- Cameron AC, Trivedi PK. 1986. Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* **1**(1): 29-53.
- Cohen WM, Levinthal DA. 1990. Absorptive capacity: a new perspective on learning and innovation. *Administrative science quarterly* **35**(1): 128-152.
- Collins BE, Guetzkow HS. 1964. *A social psychology of group processes for decision-making*. Wiley New York.
- Daft RL, Lengel RH. 1986. Organizational information requirements, media richness and structural design. *Management science* **32**(5): 554-571.
- Fleming L. 2001. Recombinant uncertainty in technological search. *Management Science* **47**(1): 117-132.
- Fleming L. 2007. Breakthroughs and the " Long Tail" of innovation. *MIT Sloan Management Review* **49**(1): 69.
- Fleming L, King Iii C, Juda AI. 2007. Small worlds and regional innovation. *Organization Science* **18**(6): 938-954.
- Freeman LC. 1979. Centrality in social networks conceptual clarification. *Social networks* **1**(3): 215-239.
- Gelman A, Hill J. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press Cambridge, MA.
- Ghosh A, Rosenkopf L. 2014. Shrouded in structure: Challenges and opportunities for a friction-based view of network research. *Organization Science*.
- Glynn MA. 1996. Innovative genius: A framework for relating individual and organizational intelligences to innovation. *Academy of management review* **21**(4): 1081-1111.
- Granovetter MS. 1973. The strength of weak ties. *American Journal of Sociology* **78**(6): 1360-1380.
- Grant RM, Jammie AP, Thomas H. 1988. Diversity, diversification, and profitability among British manufacturing companies, 1972–1984. *Academy of Management Journal* **31**(4): 771-801.
- Greene WH. 2003. *Econometric analysis*. Prentice Hall Upper Saddle River, NJ.

- Hall BH, Jaffe A, Trajtenberg M. 2005. Market value and patent citations. *RAND Journal of economics* **36**(1): 16-38.
- Hall BH, Jaffe AB, Trajtenberg M. 2001. The NBER patent citation data file: Lessons, insights and methodological tools. *NBER working paper*.
- Hall BH, MacGarvie M. 2009. The Private Value of Software Patents. *NBER working paper*.
- Hansen MT. 1999. The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative science quarterly* **44**(1): 82-85.
- Hansen MT. 2002. Knowledge networks: Explaining effective knowledge sharing in multiunit companies. *Organization science* **13**(3): 232-248.
- Hargadon A, Sutton RI. 1997. Technology brokering and innovation in a product development firm. *Administrative science quarterly*: 716-749.
- Hausman J, Bronwyn H. 1984. Econometric Models for Count Data With an Application to the Patents-R&D Relationship. *Econometrica* **52**(4): 909-938.
- Henderson RM, Clark KB. 1990. Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly* **35**(1): 9-30.
- Huber GP, Daft RL. 1987. The information environments of organizations. In *Handbook of Organization Communication*. Jablin F, Putnam LL, Roberts KH, Porter LW (eds.), Sage: Beverly Hills, CA.
- Lai R, D'Amour A, Fleming L. 2009. The careers and co-authorship networks of U.S. patent-holders, since 1975. *Working Paper*.
- Lin N. 2001. *Social Capital*. Cambridge University Press: Cambridge, MA.
- March JG, Simon HA. 1958. *Organizations*. John Wiley & Sons Inc: New York.
- McFadyen MA, Cannella AA. 2004. Social Capital and Knowledge Creation: Diminishing Returns of the Number and Strength of Exchange. *The Academy of Management Journal* **47**(5): 735-746.
- Mors ML. 2010. Innovation in a global consulting firm: when the problem is too much diversity. *Strategic Management Journal* **31**(8): 841-872.
- Nerkar A. 2003. Old is gold? The value of temporal exploration in the creation of new knowledge. *Management Science* **49**(2): 211-229.
- Nerkar A, Paruchuri S. 2005. Evolution of R&D capabilities: The role of knowledge networks within a firm. *Management Science* **51**(5): 771-785.
- Nonaka I. 1994. A dynamic theory of organizational knowledge creation. *Organization science* **5**(1): 14-37.
- Nonaka I, Toyama R, Nagata A. 2000. A firm as a knowledge-creating entity: a new perspective on the theory of the firm. *Industrial and corporate change* **9**(1): 1-20.
- Podolny JM. 2001. Networks as the Pipes and Prisms of the Market<sup>1</sup>. *American journal of sociology* **107**(1): 33-60.
- Reagans R, McEvily B. 2003. Network structure and knowledge transfer: The effects of cohesion and range. *Administrative Science Quarterly*: 240-267.
- Reagans RE, Zuckerman E, McEvily B. 2007. On Firmer Ground: The Collaborative Team as Strategic Research Site for Verifying Network-Based Social Capital Hypotheses. In *The missing links: formation and decay of economic networks*. Russell Sage Foundation.
- Rodan S, Galunic C. 2004. More than network structure: how knowledge heterogeneity influences managerial performance and innovativeness. *Strategic Management Journal* **25**(6): 541-562.
- Rosenkopf L, Nerkar A. 2001. Beyond local search: Boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal* **22**(4): 287-306.

- Schilling MA, Fang C. 2014. When hubs forget, lie, and play favorites: Interpersonal network structure, information distortion, and organizational learning. *Strategic Management Journal* **35**(7): 974-994.
- Schilling MA, Phelps CC. 2007. Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. *Management Science* **53**(7): 1113-1126.
- Singh J. 2005. Collaborative networks as determinants of knowledge diffusion patterns. *Management Science* **51**(5): 756-770.
- Singh J. 2008. Distributed R&D, cross-regional knowledge integration and quality of innovative output. *Research Policy* **37**(1): 77-96.
- Singh PV, Tan Y, Mookerjee V. 2008. Network effects: The influence of structural social capital on open source project success. *Management Information Systems Quarterly*, Forthcoming.
- Sorenson O, Rivkin JW, Fleming L. 2006. Complexity, networks and knowledge flow. *Research Policy* **35**(7): 994-1017.
- Sutton RI, Hargadon A. 1996. Brainstorming groups in context: Effectiveness in a product design firm. *Administrative Science Quarterly*: 685-718.
- Szulanski G. 1996. Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic management journal* **17**(S2): 27-43.
- Tallman S, Li J. 1996. Effects of international diversity and product diversity on the performance of multinational firms. *Academy of Management journal* **39**(1): 179-196.
- Thompson P. 2006. Patent citations and the geography of knowledge spillovers: Evidence from inventor-and examiner-added citations. *The Review of Economics and Statistics* **88**(2): 383-388.
- Tortoriello M, McEvily B, Krackhardt D. 2014. Being a Catalyst of Innovation: The Role of Knowledge Diversity and Network Closure. *Organization Science*.
- Wasserman S, Faust K. 1994. *Social network analysis: Methods and applications*. Cambridge Univ Press: Cambridge, MA.

## Tables and Figures

**FIGURE 1: SOFTWARE INVENTOR COLLABORATION NETWORK**



**Table 1: DESCRIPTIONS AND DEFINITIONS OF CONTROL VARIABLES**

Measure name	Measure Description	Logic for inclusion
Co-inventor ability	Average number of normalized forward citations for all patents applied for on which co-inventors are co-inventors.	The ability of inventors may affect both their network positions and the innovative performance of the patents they co-invent.
Co-inventor betweenness centrality	Average betweenness centrality of co-inventors, betweenness centrality for individual inventors calculated using Freeman's (1979) measure.	Co-inventors that occupy more central positions in a network can potentially control both the information flow and the resources of the network.
Repeat ties between co-inventors	Number of times co-inventors have worked together on patents during the three year moving window.	Co-inventors may work together repeatedly over time in ways that restrict their new network relationships and their access to new knowledge and their ability to innovate.
Third-party ties in co-inventor network	Number of unique common inventors with whom a pair of first time co-inventors on the focal patent have worked in the last three years.	Inventors could expand their collaboration ties in highly restricted ways that similarly limit both network structure and innovation performance.
Network size	The total number of direct and indirect ties to inventors on the focal patent.	Co-inventors that have many contacts have access to larger network resources.
Network constraint	The cohesion of the social network measured using Burt's (1992) network constraint measure.	Inventors who bridge structural holes may have informational advantages that improve their performance (Burt, 1992)
Prior art age	Average patent number for each of the patents that the focal patent cites.	Patents based on newer technology tend to have more forward citations (Nerkar, 2003)
Patent claims	Number of patent claims made by the focal patent.	The number of claims made by a patent positively correlates with the forward citation counts received by the patent (Nerkar & Paruchuri, 2005)
Technology breadth	defined for the focal patent $i$ as $1 - \sum_j s_{ij}^2$ , where $s_{ij}$ refers to the fraction of patents cited by patent $i$ that belong to technology class $j$ (Argyres & Silverman, 2004).	Patents that are based on a broad range of previous technologies will lead to solutions that are more broadly applicable and will result in innovations that have a greater and wider impact (Rosenkopf & Nerkar, 2001).
Number of patent citations	Number of citations made by the focal patent	Patents that cite more patents are more likely to appear in the search results of both examiners and other inventors, resulting in more forward citations.
Number of non-patent references	Number of citations to academic or theoretical knowledge rather than other patents	Patents that build on academic or theoretical knowledge are recognized as fundamental innovations by inventors and therefore receive more citations (Nerkar & Paruchuri, 2005).
Technology class	Dummy variable indicating the primary three digit technical class of the focal patent.	Patents belonging to different technical categories differ in the number of citations received (Hall <i>et al.</i> , 2001).

Regional performance	Number of patents applied for in the region of the focal patent in the prior three years. Address of first inventor used (Singh, 2005, 2008; Thompson, 2006) to determine MSA (Singh, 2008; Thompson, 2006)	Patents that originate in certain geographic regions may receive more citations than patents in other regions.
No backward citation dummy	Dummy variable takes a value of 1 if the focal patent has no backward citations and 0 otherwise.	Some of the patents have no backward citations, so many of the patent attributes described above cannot be calculated for those patents.
Number of co-inventors	Total number of inventors that collaborated on a patent	Patents that have large number of co-inventors have more resources available.

**TABLE 2: DESCRIPTIVE STATISTICS AND TRANSFORMATION OF KEY VARIABLES**

Variable Number	Variable Name	Number of Observations	Mean	Standard Deviation	Min	Max	Transformation
1	Forward citations	50066	23.7695	29.2389	0	559	NA
2	Value of direct contacts' combinatory knowledge	50066	0.076	0.2183	0	1	Log transformation
3	Direct contacts' knowledge heterogeneity	50066	.0873	0.2217	0	0.9135	Log transformation
4	Value of indirect contacts' combinatory knowledge	50066	0.0496	0.1594	0	1	Log transformation
5	Indirect contacts' knowledge heterogeneity	50066	0.119	0.2659	0	0.9326	Log transformation
6	Prior art age	50066	4954221	681893	0	6174603	Square transformation and scaled down by $10^{13}$
7	Number of claims	50066	18.4910	15.0907	1	375	Log transformation
8	Number of non-patent references	50066	2.9320	10.8345	0	394	Log transformation
9	Number of patent backward citations	50066	11.3731	17.3709	0	352	Log transformation
10	Technical breadth of patents	50066	0.4892	0.2634	0	0.93	Square transformation
11	Regional performance	50066	1170.0930	1425.2930	0	4676	Log transformation
12	Number of contacts	50066	2.302	5.7871	0	112	Log transformation
13	Third party ties	50066	0.0398	0.4829	0	33	Log transformation
14	Repeat ties	50066	1.6276	11.4359	0	786.96	Log transformation
15	Network Constraint	50066	0.1281	0.2328	0	1.12	Log transformation
16	Betweenness centrality	50066	0.0006	0.0080	0	0.33	Log transformation
17	Co-inventor ability	50066	15.1039	26.1593	0	559	Log transformation
18	Number of co-inventors	50066	2.565	1.8726	1	32	Log transformation

**TABLE 3: CORRELATION MATRIX**

(The numbers correspond to the variable numbers in Table 2)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	1																		
2	0.04	1																	
3	0.03	-0.07	1																
4	0.01	0.10	-0.01	1															
5	0.03	-0.06	0.35	-0.03	1														
6	-0.09	-0.02	0.16	0.00	0.18	1													
7	0.10	0.01	0.09	0.02	0.09	0.22	1												
8	0.11	0.01	0.06	0.01	0.07	0.06	0.13	1											
9	0.06	0.01	0.05	0.01	0.06	0.23	0.09	0.15	1										
10	0.06	0.02	0.02	0.01	0.02	0.05	0.05	0.10	0.45	1									
11	-0.11	-0.03	0.15	0.00	0.16	0.36	0.11	0.04	0.00	-0.04	1								
12	0.04	0.04	0.34	0.09	0.35	0.17	0.09	0.06	0.07	0.01	0.17	1							
13	0.01	-0.02	0.20	0.01	0.21	0.04	0.03	0.01	0.02	0.00	0.03	0.20	1						
14	0.00	0.03	0.28	0.04	0.31	0.02	0.01	0.08	0.09	0.03	0.08	0.33	0.14	1					
15	-0.02	0.24	0.21	0.11	0.25	0.09	0.04	0.01	0.02	-0.02	0.09	0.31	0.08	0.05	1				
16	-0.02	-0.02	0.25	-0.01	0.25	0.07	0.03	0.02	0.02	-0.01	0.07	0.26	0.07	0.12	0.03	1			
17	0.03	0.23	0.40	0.12	0.36	0.13	0.09	0.09	0.10	0.03	0.14	0.33	0.13	0.45	0.53	0.10	1		
18	0.09	0.04	0.19	0.04	0.21	0.00	0.04	0.07	0.08	0.06	0.00	0.13	0.16	0.41	-0.08	0.03	0.25	1	

**TABLE 4: COMPARISON OF MEANS BASED ON SAMPLE CUTS FOR HIGH AND NON-HIGH VALUES OF THE INDEPENDENT VARIABLES**

Variable based on which group is constructed	Mean normalized forward citation for high group	Mean normalized forward citation for non-high group	% difference between high and non-high groups
Direct contacts' combinatory knowledge	1.12	0.87	22.32%
Direct contacts' knowledge heterogeneity	0.95	0.89	6.7%
Indirect contacts' combinatory knowledge	0.96	0.89	7.8%
Indirect contacts' knowledge heterogeneity	1.04	0.89	16.85%

**TABLE 5: REGRESSION RESULTS OF NEGATIVE BINOMIAL MODELS WITH INNOVATIVE PERFORMANCE AS DV**

	Model 1 - Constant Only	Model 2 – value of combinatory knowledge of direct and indirect contacts	Model 3- knowledge heterogeneity of direct and indirect contacts	Model 4 - Both value of combinatory knowledge and knowledge heterogeneity of direct and indirect contacts
	Forward citations	Forward citations	Forward citations	Forward citations
Value of direct contacts' combinatory knowledge			0.1259*** (0.0388)	0.1319*** (0.0388)
Value of indirect contacts' combinatory knowledge			0.0166 (0.0594)	0.0429 (0.0601)
Direct contacts' knowledge heterogeneity		-0.0004 (0.0517)		0.0015 (0.0518)
Indirect contacts' knowledge heterogeneity		0.1112** (0.0527)		0.1227** (0.0531)
Prior art age	0.3027*** (0.0289)	0.3018*** (0.0289)	0.3039*** (0.0289)	0.3028*** (0.0289)
Number of claims	0.1308*** (0.0062)	0.1307*** (0.0062)	0.1307*** (0.0062)	0.1306*** (0.0062)
Number of patent backward citations	0.0395*** (0.0068)	0.0394*** (0.0068)	0.0397*** (0.0068)	0.0397*** (0.0068)
Number of non-patent references	0.0249*** (0.0018)	0.0249*** (0.0018)	0.0249*** (0.0018)	0.0249*** (0.0018)
Technology Breadth	0.1351*** (0.0254)	0.1355*** (0.0254)	0.1335*** (0.0254)	0.1337*** (0.0254)
No backward citation dummy	0.6742*** (0.0789)	0.6729*** (0.0789)	0.6776*** (0.0789)	0.6762*** (0.0789)
Regional performance	0.0038 (0.0037)	0.0037 (0.0037)	0.004 (0.0037)	0.0039 (0.0037)
Number of contacts	-0.0088 (0.086)	-0.0313 (0.127)	-0.0024 (0.088)	-0.272 (0.128)
Third party ties	-0.0043 (0.0063)	-0.0053 (0.0063)	-0.0037 (0.0063)	-0.0048 (0.0063)
Repeat ties	-0.0108*** (0.0025)	-0.0110*** (0.0025)	-0.0106*** (0.0025)	-0.0108*** (0.0025)
Network Constraint	-0.0237*** (0.0050)	-0.0207*** (0.0051)	-0.0276*** (0.0051)	-0.0246*** (0.0052)
Betweenness centrality	0.0206 (0.0279)	0.0205 (0.0279)	0.0179 (0.0280)	0.0178 (0.0280)
Co-inventor ability	0.0328*** (0.0033)	0.0331*** (0.0034)	0.0313*** (0.0034)	0.0316*** (0.0034)
Number of co-inventors	0.1223*** (0.0083)	0.1180*** (0.0085)	0.1227*** (0.0083)	0.1178*** (0.0085)
Constant	1.4620*** (0.1483)	1.4728*** (0.1485)	1.4289*** (0.1488)	1.4397*** (0.1490)
Number of observations	50066	50066	50066	50066
Log likelihood	-200206.7334	-200202.0593	-200199.9929	-200194.3317

Note: Robust standard errors are in parentheses; \*\*\*p<0.01, \*\*p<0.05, \*p<0.1; firm, application year and three-digit technical class dummies are omitted from the table.