The Impact of Networks and Technological Specialization on the Value of Academic Entrepreneurs

Pankaj Patel
Ball State University

Phillip Phan
Johns Hopkins University

ABSTRACT

In what ways do technology ventures in government funded programs gain from increased academic inventor involvement? The literature finds mixed support for the role of academic entrepreneurs (AE) in providing scientific and tacit knowledge to start up technology ventures. To resolve these findings, we posit that an AE's contribution to start-up success depends on (a) the extent of 'shared specialization', the degree to which the scientific domain of an innovation overlaps with that of the AE's expertise, and (b) the structure of the AE's innovation network. Shared specialization measures the continuity of an innovation along a technological trajectory and the extent of its overlap with related trajectories. Using data on the network structure of scientists involved in biotech ventures, we find that AEs are more valuable to ventures when shared specialization is high, and when their network structures display high levels of centrality, structural holes, and knowledge combinability. The findings have implications for public policies related to entrepreneurship. Specifically, encouraging or requiring shared specialization between ventures and academic entrepreneurs might improve the success of such programs as the Small Business Innovation Research (SBIR).

Key words: Academic entrepreneurship, shared specialization, bibliographic coupling, knowledge combinability, knowledge structure

Under Journal Review
Do not quote without permission

---

1. Introduction

While much of the research on venture funding has focused on private sector funding for ventures from VCs and angel investors, public sector programs related to the Federal Government's Small Business Innovative Research (SBIR) program are under increased scrutiny (Lerner, 1999). Central to the mixed findings on the effectiveness of such programs as the SBIR could be the underlying nature of the businesses that receive such funding. Started with the intention of funding early stage technology development in small firms, the SBIR program now represents 60 percent of all public financed SME programs, and is two-thirds the size of private venture capital funds (Audretsch, 2003). SBIR grants were developed to address gaps in funding markets that were not filled by private venture capital. But the

research has shown that in those industries where VC funding is low, SBIR funded projects are more prone to failure (Gans *et al.*, 2000; Lerner, 1999). This result calls into question the viability and utility of SBIR grants in providing support to businesses in a broad range of technological and geographic areas. With 16,510 SBIR awards representing $5.308 billion dollars[1] in funding between 2001 and 2010, improving the success of SBIR awards across diverse technological and geographic areas is central to the viability of its initial Congressional mandate that "each federal agency [must allocate] around four percent of its annual budget to funding innovative small firms as a mechanism for restoring American international competitiveness" (Audretsch, 2003, page 132).

Given that the SBIR program seeks to promote applied research in basic sciences through a private-public partnership arrangement, it stands to reason that an assessment of the strategic outcomes of the program, in the context of the resource base, the academic researchers, underpinning the projects would provide useful empirical lessons for how such programs and the firms' responses to them could be better structured. This research answers a simple question, 'In what ways do technology ventures that are publically funded gain from increased academic inventor involvement?' Put another way, 'In the light of the social goods that emanate from the public funding for basic research, should the public funding for applied research only occur if the academic inventor is involved?'

The importance of the SBIR program in developing ventures from basic science is evident from several prior findings on SBIR funded ventures. First, there is significant partnership between private and public sector (e.g. universities) in filing for SBIR grants (Audretsch *et al.*, 2002a). As universities play a fundamental role in the advancement of the basic sciences, the involvement of university researchers in SBIR ventures could be central to their success. Similarly, SBIR awards have been central to helping scientists shift from basic research to entrepreneurship (Audretsch *et al.*, 2002b; Bessette, 2003; Feldman, 2000; Toole *et al.*, 2007b). In sum, we know that the link between SBIR awards and involvement of academics from the basic sciences is strong. We know less if this link produces better innovation outcomes.

The literature tells us that inventor involvement enhances the stock of scientific knowledge in a venture (Agrawal, 2006) and provides tacit knowledge to help identify new technological opportunities. Some researchers have argued that an academic entrepreneur's involvement increases the information gathering intensity, problem solving ability, patent productivity, and commercialization success of a venture (Cockburn *et al.*, 1998; Shane, 2004; Toole *et al.*, 2009; Zucker *et al.*, 1996; Zucker *et al.*, 2002; Zucker *et al.*, 1998).

1. report.nih.gov/FileLink.aspx?rid=535

Under Journal Review
Do not quote without permission

There are also circumstances in which an inventor's knowledge has limited applicability to the commercialization efforts of a venture (Gittelman *et al.*, 2003; Stern, 2004). Academic inventors, for example, possess institution-specific human capital that allows them to operate under academic norms more effectively than in commercial settings. When such inventors become entrepreneurs, their human capital may limit the "degrees of freedom" needed by ventures to develop the requisite knowledge base from which future innovations and opportunities can emerge. For example, academic entrepreneurs may be less sensitive to the need for market or competitive knowledge, which can limit the degree to which they are willing to invest in the acquisition of such information.

Academic knowledge capital is highly tacit, stemming from the trial-and-error process of discovery and invention (Agrawal, 2006). Academic discoveries tend to start from basic science, since these lines of inquiry are more likely to lead to novel and publishable findings. While based on theoretical first principles, this 'hypothesis testing' process is iterative and, for truly novel discoveries, can include knowledge from distantly related scientific domains of inquiry, which further renders the outcomes highly uncertain[2]. Similarly, the progression from basic science to applied science to translational science is iterative, with elements of trial and error, involving the participation of users and researchers, and can take unexpected directions[3]. The earlier stages of this process are generally not replicated by commercial research enterprises due to the increasingly stringent requirements to show early and measurable returns on investment in research and development.

Academicians tend to focus on narrow scientific domains in their research since success in publishing depends on appeasing the gatekeepers of a domain with incentives to define the domain specifically rather than generally. The institutional specificity of an AE's knowledge (scientific history of a discovery, appropriate funding programs, scientifically significant questions and finding, and the norms of research team dynamics) may be less valuable in commercial contexts, especially if the set of scientific domains representing a venture's technological offering is broad and heterogeneous. Broad technological domains require information to be accumulated from extensive search activities so as to foster new knowledge combinations. Extensive searches are costly and therefore consume resources that are critical to a new venture's survival.

Given the above discussion it stands to reason that the knowledge capital of AEs in narrower technological domains would be more valuable. Because of extensive experience with trial and error searching in their research work, AEs are able to more able to conduct optimal local searches

[2] For example, the combination of chemistry and computer visualization (computational chemistry) in rational drug design resulted in the creation of new classes of selective serotonin reuptake inhibitors for the treatment of depression.

[3] The classic example of this is Pfizer's sildenafil citrate (Viagra), which was originally designed to treat pulmonary hypertension.

Under Journal Review
Do not quote without permission

productively. They can more easily access the knowledge bases represented by their areas of expertise and to recognize scientific problems and breakthroughs. For example, AEs are networked to academic associations that routine share highly specialized knowledge and. Following Von Wartburg et al. (2005), we posit that when the *shared specialization* of an underlying innovation is high, AEs are more likely to be central to innovation success.

Shared specialization refers to the technological genealogy of an innovation and is given by two dimensions, technological specialization and technological sharedness. The former describes how closely an innovation adheres to its scientific genealogy while the latter describes how closely a cluster of innovations overlap in terms of their underlying scientific domains.

Technological specialization and technological sharedness are related constructs. An innovation evolving along a scientific trajectory may not continue to share its technological domain with prior innovations after a certain point of development. For example, the technological paradigm in the solar energy industry is the extraction of energy by the conversion of photons to electrons. As applications in solar technology diverge, technological sharedness declines over time. This can occur when the value added from complementary technologies, for example, photovoltaic foils (photosensitive thin film), photovoltaic paints (photosensitive pigments), and photovoltaic tiles (photosensitive construction material), increases faster than the value created by the core technology.

Other innovations continue to rely on the same basic technology as they mature in design, performance, and applications. In these cases, the degree of technological sharedness increases as the core technology becomes more valuable with the addition of complementary technologies. For example, light emitting diode (LED) technology has not changed despite the development of new applications along different technological trajectories ranging from infra-red to ultraviolet light spectrum devices. As new applications for LED devices are found, such the evolution from ambient to task to space lighting, the value of the technology and its accompanying investments increases over time.

Shared specialization is particularly relevant for our discussion of the role of AEs in venturing success for several reasons. Drawing on notions from Kauffman's *NK* dynamic system model, shared specialization is anticipated to increase the effectiveness of local knowledge search (Kauffman *et al.*, 1989). When shared specialization is high, the configurations of the number of technological elements in a product or service are fewer (i.e. fewer *N*) and they exhibit greater interdependency (i.e. higher *K*). AEs that can manage higher *Ks* by deploying their specialized knowledge across fewer *Ns* render local searches more effective.

2. Conceptual Framework and Hypotheses

Under Journal Review
Do not quote without permission

Ventures generate rents by leveraging human capital to combine knowledge in novel ways; the success of which depends on the breadth and depth of internal knowledge stocks concomitant with the acquisition and incorporation of external knowledge. The challenge for ventures in technology commercialization is therefore to access a diverse set of knowledge while at the same time find ways to productively combine them with internal knowledge bases (Dushnitsky *et al.*, 2005; Shane, 2000). The ease of combining knowledge is in turn contingent on the complexity of the technological domain in which the venture is located. Knowledge combinations are easier when ventures focus on narrower and deeper knowledge domains that are similar to those of the firm's since existing personnel can more readily appreciate the value and use of such knowledge. Trade off is that increased but narrower knowledge depth is more likely to result only in incremental innovation since the possibilities of truly novel combinations that lead to radical innovation are less (Laursen *et al.*, 2006).

New ventures face the liability of newness (Stinchcombe, 1965), which means that they have to focus resources on a narrow set of priorities and knowledge domains to overcome liability threats. Recent work by Kaplan, Sensoy and Strömberg (2009), suggests that ventures are less likely to radically change the underlying technology of a product over its life time. Unlike larger firms that can radically renew its technological knowledge by scanning across more technology elements (*N*) at higher levels of complexity (*K*), ventures lack resources to engage in such searches.

Furthermore, from the perspective of stakeholders ventures focusing on innovations that require broad knowledge searches may be view as less legitimate (garnering lower levels of support from resource providers such as venture capitalists) since broad search imply the expense of valuable resources on developing new sets of symbols, language, and exchange routines rather than on the execution of business priorities.

Ventures can more easily realize rents by focusing on narrow technological domains, and engaging in deeper and intense searches in the narrowed domain. By increasing the degree of overlap among narrow sets of knowledge, ventures can engage in more incremental experimentation while more efficiently manage scarce resources. The deeper searches into overlapped knowledge bases are also useful for two reasons. First, because of the liability of newness, ventures may lack knowledge routines and processes to engage in complex searches across many knowledge domains. Second, because of the liability of smallness, ventures with the requisite knowledge routines are less likely to enjoy economies of scale and scope in exploiting knowledge acquired through broad searches.

To ultimately succeed, ventures must commercialize their discoveries. In this context, the notion of shared specialization is particularly relevant. Recall that technological specialization focuses on the continuity of the technological genealogy of an innovation. Ventures focusing on innovation with increased technological specialization can more easily acquire resources from the task environment as

Under Journal Review
Do not quote without permission

stakeholders perceive lower uncertainty with respect to the odds of commercialization. A venture with a history of successful commercialization from its discovery efforts, it can draw on existing symbolic and institutional frameworks to establish boundaries, acquire resources, and develop exchange relationships with future stakeholders.

2.1. AE Involvement in Ventures

According to our earlier discussion, firms increase their innovative capacity by local and distant information searches. Although distant searches lead to breakthrough innovations, they are costlier than local searches, and are more risky. Local searches cost less and are less risky but more likely to result in incremental innovation. Due to financial exigencies, start ups are therefore more likely to rely on local searches to innovate because survival imperatives trump the need to make radical change early in the life of a venture (Stuart *et al.*, 1996). Entrepreneurs who are not themselves scientists have no capability to conduct distant scientific searches and so must rely on pre-existing knowledge stocks within the start-up or they must acquire the capability by hiring or empanelling scientific experts. An AE's involvement provides a venture with the scientific knowledge to conduct distant searches that, when combined with existing knowledge stocks, are more likely to lead to breakthrough innovations.

Zucker et al. (2002) and Cockburn and Henderson (1998) propose that increased AE involvement in a venture enhances knowledge combinability by facilitating the transfer of tacit human capital in an organization. AEs have intimate knowledge of the underlying science and the trial and error process of the invention (Agrawal, 2006). Others have further suggested that an AE's ability to facilitate the transfer of knowledge is institutionally specific and hence may actually constrain the venture's commercialization efforts. Because academic research is more oriented toward the basic sciences, it is possible that the commercial applications of such knowledge are limited (Stern, 2004).

The knowledge structuration perspective says that when ventures draw from narrow technological domains they must delve deeper into their knowledge bases to identify new innovation opportunities (Berger *et al.*, 1967). An AE's institutional knowledge is tuned to exploration in the basic sciences (Toole *et al.*, 2009). Therefore, her involvement in ventures with narrow technological domains creates more opportunities for the variation, selection, and retention of innovations (Kogut, 2000). Additionally, the tacit knowledge that accompanies such competencies is more transferable in narrow technological domains than broad ones.

Conversely, the acquisition and transfer of knowledge is more challenging for innovations with less shared specialization because of the increased distance between the AE's knowledge base and that required for successful commercialization in the venture (Nooteboom, 2000).

Under Journal Review
Do not quote without permission

Hypothesis 1: AE involvement in innovations with higher levels of shared specialization in technology would have a positive effect on venture performance.

2.2. Knowledge Combinability

The social constructivism theory of knowledge suggests that individuals in different environments interpret, understand, and evaluate the world differently (Berger *et al.*, 1967). Because of the stability of institutional forms in academic settings, AEs have relatively similar institutional knowledge. Therefore, the 'cognitive distance' among different types of scientific domains is relatively small and more easily combined (Nooteboom, 2000)[4]. Knowledge combinability thus refers to the ability of an AE to seek and exploit knowledge from diverse scientific domains.

Knowledge combinability is therefore a particularly important consideration for ventures in narrow technological domains that need to find new innovation pathways. Without the introduction of new knowledge, such ventures are limited to experimenting with the levels of interdependencies within the elements in the domain. When knowledge combinability is high, AEs can more easily integrate distant scientific domains into the venture's existing knowledge base and so increase the productivity of innovation attempts.

Hypothesis 2: Knowledge combinability positively moderates the relationship between shared specialization and venture performance.

2.3. Network Centrality

Consideration for the knowledge characteristics of innovations is insufficient to explain why AEs may be more valuable in some ventures and not others. Uncertainty in the outcomes of innovation attempts can create new knowledge demands. When making decisions under uncertainty, the ability to *access* diverse information is important for reducing that uncertainty (Leblebici *et al.*, 1981). While an AE's institutional knowledge and ability to combine knowledge are important drivers of innovation, the knowledge derived from her network also matters. The social network of an AE holds the pool of information from which she can introduce new knowledge to the venture so that the diversity of such a knowledge network represents the boundaries within which an AE can search. Contact with scientists in

Under Journal Review
Do not quote without permission

[4] Examples are myriad and include computational chemistry, biomechanical engineering, neurophysics, public health marketing, and so on. Virtually any combination of the natural and social sciences is being explored as academic departments seek distinction in research contribution to their fields of inquiry.

an AE's knowledge network leads to the exchange of tacit knowledge, especially with those individuals that have worked closely with AEs in the past.

The basic dimensions of a social network is given by the relative location of an actor to the part of the network with the densest connections (centrality), and the degree to which the actor has connections that are not replicated elsewhere in the network (structural hole). Ibarra (1993) shows that the centrality of an actor increases innovation in the organization. This is because occupying a central position in a network makes it easier for an actor to access knowledge held in the network because of the density of direct connections she enjoys with others (Bonacich, 1987; Brass *et al.*, 1993). As earlier discussed, ventures operating in knowledge domains with higher shared specialization require expansive searches to find novel discoveries. Centrality renders an AE's distant knowledge search more productive because it reduces the information overload from inefficient searches (Gilsing *et al.*, 2008).

Hypothesis 3: Network centrality positively moderates the relationship between shred specialization and venture performance.

2.4. Structural Holes

Together with the importance of network centrality to efficient information search, an AE's structural holes also make a difference to her contributions to the venture. A structural hole is defined as a place in a network in which two actors are indirectly connected via a third actor. That third actor is said to occupy a *structural hole* and is therefore a gatekeeper of the information flow between the unconnected actors. Therefore, structural holes provide access to the unique (non-redundant) information held in a network (Burt, 1995). An AE that spans many structural holes increase the span of her scientific domain and can combine knowledge in ways that other actors in the network cannot because they do not have similar information stores. Conversely, an AE that is embedded in a dense network (i.e., one in which every actor is connected directly to every other actor) does not have an advantage since everyone has the same information. The ability to create new knowledge and make novel discoveries is therefore attenuated when an AE is part of a dense network.

An AE in a network rich with structural holes can search distant scientific domains more efficiently and quickly because she does not have to bear the additional costs of sifting, storing, and processing redundant information held by other actors in the network. This, in turn, increases the productivity of local searches in situations where shared specialization is high because the AE has access to a store of unique external knowledge that she can combine with venture's technological knowledge (Fleming *et al.*, 2004; Yayavaram *et al.*, 2008).

Hypothesis 4: Network structural holes positively moderate the relationship between shred specialization and venture performance.

Under Journal Review
Do not quote without permission

3. Data and Measures

The empirical test of our theoretical model employs data collected from biotechnology ventures that participated in the U.S. Small Business Innovation Research (SBIR) funding program between 1996 and 2005[5]. The academic entrepreneurs are university-affiliated principal investigators (PIs) that led the commercialization efforts of the ventures funded by SBIR grants. Most of the ventures spin offs from University disclosures.

Following Toole and Czarnitzki (2007a, 2009) we use the SBIR Phase 1 award as a sampling frame so that all the ventures in our sample received SBIR Phase 1 awards. We identified 6793 unique PIs receiving 42,688 SBIR awards made between 1996 and 2005.[6] We then match the names of the PIs with the National Institute of Health (NIH) database for recipients of grants from the NIH. Specifically, we use the NIH Research Portfolio Online Reporting Tools (RePORT)[7] to match the names of SBIR PIs with NIH grant recipients.

Using the approach suggested by Toole *et al.* (2009) individuals who received NIH grants *before* they received the SBIR grants are categorized as academic entrepreneurs (i.e., they were first research academics before being entrepreneurs). We then matched the names from the SBIR and NIH databases[8] and identified 376 academic entrepreneurs. We then manually checked the list to ensure consistency. Finally, to confirm that an AE is an active biomedical researcher, we cross-referenced their names with the *PubMed* database. No discrepancies were found.

We then matched the 376 AEs with patents in the United States Patent and Trademark Office (USPTO) database. Of the 376 AEs, 22 had no matching records in the patent database. It is likely that patents were not identified due to name matching errors. We therefore manually searched the curriculum vitas of the 22 AEs and were able to construct a complete list of patent records for 14 of the 22 AEs. We dropped the remaining eight AEs. Seventeen ventures had more than one AE so to ensure that our data is consistent with our hypotheses on *individual* inventors, we drop the 17 ventures from further analyses. The final sample consisted of 357 AEs.

Under Journal Review
Do not quote without permission

[5] Each year, Federal agencies with extramural R&D exceeding $100 million are required to allocate 2.5% of their R&D budgets to the SBIR and STTR programs. Currently, the Departments of Agriculture, Commerce (NIST and NOAA), Defense, Education, Energy, Health and Human Services, Homeland Security, and Transportation, EPA, NASA, and NSF participate in these programs.

[6] Small Business Advocacy (SBA) maintains award information under the publicly available SBA TECH-Net (U.S. Small Business Administration Technology Resources Network) database. It provides the full name of the PI, location of the PI, title of the project, phase of the award (Phase I or II), award amount in each phase, company name, and whether the firm is minority or female owned (http://web.sba.gov/tech-net)

[7] http://www.projectreporter.nih.gov/reporter.cfm

[8] Using SRA NetOwl

Then, using a genetic matching algorithm, we developed a matched set of SBIR-funded ventures without AE involvement to assess the degree to which AE involvement increased a venture's chances of receiving a Phase II SBIR follow-up award and the breadth and depth of patent citations of its technology after the initial funding event.

Using the company name in the SBIR award database, we collected additional data on venture capital funding from SDC VentureXpert, and firm size data from Dunn and Bradstreet, which was crossed checked with the corresponding SBIR information.

3.1. Independent Variables

This study employs two sets of independent variables, shared specialization and AE knowledge measures. Earlier research indicates that an invention patent cannot be a proxy for an innovation since an innovation can contain many complementary patents. However, in the context of the current sampling frame, patents provide a reliable and valid measure of shared specialization because the SBIR grant program's main objective is to convert an invention or 'proof-of-concept' into a viable innovation. The underlying principal patent in the SBIR therefore proxies for an innovation as the principal investigator and the venture team have made significant time and monetary investments.

3.1.1. Shared Specialization.

Following von Wartburg et al. (2005), we defined shared specialization as the mean of two knowledge network measures: (i) reachability out-degree (Figure A.1.1, Appendix 1) (ii) bibliographical coupling (Figure A.1.2, Appendix 1). Von Wartburg et al. (2005), in order to conceptualize their measure of shared specialization traced the invention and evolution of the valve actuator mechanism in the internal combustion engine. By using a core invention, the lineages of future inventions could be easily tracked. We cannot follow this technique in the current research because we have the opposite situation. We do not know the core patent family from which the venture invention originated but we can identify the underlying patent supporting the SBIR application.

To identify the core patent family of the underlying innovation, we created patent landscapes of the network of cross-cited patents. We use the six-degrees of separation standard to construct the patent landscape around the focal patent for which the SBIR grant is based. Six-degrees of separation has been widely used in network analysis to identify comprehensive sets of network nodes. For the combined matched pair (i.e., with and without AE involvement) sample of 714 ventures, we identified 1,972,417 patents within six-degrees of separation from the focal patent. We then develop a network of patent families for each invention.

Under Journal Review
Do not quote without permission

In the next step, we calculate the measures of reachability out-degree and bibliographical coupling for the innovation. Reachability out-degree is defined as the probability weighted Freeman out-degrees of cited patents. The degree of bibliographical coupling is one divided by the average of the direct out-degrees of the bibliographically coupled patents. Detailed explanations of the operationalizations are available in von Wartburg et al. (2005).

Reachability out-degree also measures patent lineages in addition to direct linkages, whereas bibliographic coupling measures the extent of sharedness with a core patent family. In other words, reachability out-degree explains the level of specialization along different lineages of innovations emanating from a core patent family. Bibliographic coupling measures the extent to which the underlying invention employs a technology that is common to other innovations from the core patent family. Figure A.1.3. shows that ventures with AE involvement have higher levels of shared specialization.

3.1.2. Knowledge Combinability.

Knowledge combinability is a proxy for how well an AE is able to combine diverse sets of knowledge. Ventures face liabilities of newness and smallness and they must be able to aggregate diverse knowledge to develop the appropriate responses to environmental threats. AEs that are conversant with a broad range of knowledge domains are more able to combine diverse knowledge sets more effectively. To operationalize this measure, we took an inventory of all the patents associated with an AE ten years before the SBIR application date. We then listed all the technology three-digit[9] sub-classes that were cited in the SBIR application related patent. The measure of knowledge combinability refers to the extent to which an inventor combines divergent knowledge in her patent portfolio. Although prior measures rely on technological distance or technological similarity measures, these measures do not account for: (i) intensity of combination efforts; (ii) difficulty of distant combination efforts. In other words, the extent to which a technology class is combined across the remaining technology classes (or, intensity), and citations from distant technology classes could further increase complexity of combination.

We start by identifying developing a matrix of technology classes represented in an inventor's portfolio (T) and the citations in the inventor patents (C) across different technology classes (T). The resulting $T \times C$ shows the (i) expanse of combination efforts across T technology classes; (ii) intensity of combination efforts through citation across classes. To start, we consider each technology class as the

[9] Not all patents provided a fine grained level of detail on technology classes. As IPC-6, or an alternative classification approach, could provide a more fine-grained comparison, we chose the 3-digit classification approach. This approach is more inclusive and does not bias the measure towards Type 1 errors (inflating knowledge combination capability) or Type 2 errors (overlooking ability of broader technology classes in affecting knowledge combination in hierarchically lower technological domains).

Under Journal Review
Do not quote without permission

focal technology class (T_x), and calculate cosine similarity across the remaining *T-1* technology classes. The measure of cosine similarity is between the focal technology class (*x*) and class (*i*):

$$S_{i,x} = \frac{Class_i \cdot Class_x}{\|Class_i\|\|Class_x\|} = \frac{\sum_c^C Class_{ic} \times Class_{xc}}{\sqrt{\sum_c^C Class_{ic}^2} \times \sqrt{\sum_c^C Class_{xc}^2}}$$

Next, we weight similarity index based on Euclidean Distance (*D*) between class *i* and class *x*.

For N technology classes represented in an inventor patent portfolio, the measure of knowledge combinability is given as

$$Knowledge\ Combinability = \frac{\sum_{i,x}^N S_{i,x} \times (|D_i - D_x|)}{N}$$

Consider for example, an inventor with the following matrix

	Patent 1	Patent 2	Patent 3	Patent 4	Patent 5	Patent 6
Technology Classes 1 (#302)	3	0	2	1	4	5
Technology Classes 2 (#304)	1	3	3	4	2	1
Technology Classes 3 (#107)	1	0	0	2	3	1
Technology Classes 4 (#119)	1	7	3	2	4	6
Technology Classes x (#136)	1	2	3	1	1	4

$$S_{1,x} = \frac{Class_1 \cdot Class_x}{\|Class_1\|\|Class_x\|} = \frac{(3\times1)+(0\times2)+(2\times3)+(1\times1)+(3\times1)+(4\times1)+(5\times4)}{\sqrt{9+0+4+1+16+25} \times \sqrt{1+4+9+1+1+16}}$$

$$= \frac{3+0+6+1+3+4+20}{\sqrt{55}\times\sqrt{32}} = \frac{37}{41.92} = 0.881$$

Adjusting this for technology distance, $S_{1,x}(|D_i - D_x|) = S_{1,x}|302 - 136| = 146.246$

Similar steps are repeated for the remaining dyads of technological classes, and the final sum is rationalized by total number of technology classes.

3.1.3. Inventor Centrality.

The structural centrality of an AE network is measured by the network of other co-patentees in the AE network. We used patent information from ten-years before the venture received the SBIR Phase I award. By measuring the pre-venture knowledge network of the AE, we can know the condition of the latter's accessibility to diverse sets of knowledge. The AE knowledge network is defined as a single mode inventor-inventor collaboration network in UCINET VI. The nodes are other inventors that the AE has collaborated with in the past and the links are the co-patenting ties. Ten-years provides a reasonable time window to fully measure the academic ties of the AE.

Under Journal Review
Do not quote without permission

We use the structural centrality measure of Bonacich (1987) because it accounts for the centrality of other inventors connected to AE. Thus, the measure represents the global centrality of an AE:

$$c(\alpha,\beta) = \alpha \sum_{K=1}^{\infty} \beta^k R^{k+1}\, 1_i$$

Where $c(\alpha,\beta)$ is a vector of centrality scores for inventors, and α is an arbitrary scaling factor and β is a weight, and 1 is a column vector of ones. When $\beta > 0$, the AE who is linked to highly central others is also highly central (Bonacich, 1987). Following earlier research, we set β equal to the 3/4 of the reciprocal of the largest eigenvalue of *X* (Bonacich, 1987).

3.1.4. Structural Holes.

We use the efficiency measure of structural holes proposed by Burt (1995). The efficiency based measure uses the ratio of non-redundant contacts to total contacts for a focal inventor:

$$\left[\sum_j \left(1 - \sum_q p_{iq}\, m_{jq}\right)\right] / C_j$$

where p_{iq} is the proportion of inventor *i*'s co-patenting with contact *q*, and m_{jq} is the marginal strength (i.e. number of patents) of the relationship between contact *j* and contact *q*, and *Cj* is the total number of contacts for inventor *i*. A high value indicates that the AE has an ego network rich in structural holes (i.e., the proportion of non-redundant to redundant ties is high). For AEs without any co-inventors, the index was set to 0.1 (Ahuja, 2000).

3.2. Control Variables

External Legitimacy is the total number of alliances with established pharmaceutical firms. We use a simple count measure with data gathered from *www.bio.org* and *MedAd News*. *Location legitimacy* is the count of biotech firms in the county; we obtained data from the SBIR Phase I award reports and confirmed it with Dunn and Bradstreet. *Patent stock* the is the number of patents held by the venture prior to the SBIR application date. *Firm Size* is the number of employees at the date of SBIR application. We control for *SBIR Phase I award amount*, whether the venture is *VC backed*, and if the venture is *minority owned* or *female owned*. *Competitive intensity* is the count of the number of players represented in the 17 technology classes in the data. Finally, we include *10 time dummies* (1996-2005) and *17 technology dummies* to account for any unobservable fixed effects.

4. Dependent Variables and Estimation Model

Under Journal Review
Do not quote without permission

The dependent variables are the proxies for AE contributions to the performance of the venture. We use the amount of a Phase II Award and the forward citations of patents approved after the AE joined the venture. The Phase II Award amount is a dollar value indicating the magnitude of firm success at completing academic style proof-of-concept studies. Therefore, the SBIR reviewers judge the level of innovation performance from a scientific standpoint. Based on the performance of the venture after receiving a Phase I award, the reviewers in the NIH SBIR committees determine whether the venture can be awarded a Phase II and for how much. Therefore, this outcome variable is left censored. We also use patent forward citations to measure the firm's success at innovation oriented tasks. Numerous studies have used forward citation as an indicator of success in knowledge development. Forward citation is a censored count variable.

4.1. Genetic Matching.

The sample of AE involved ventures represents 5.25% (357 AE-ventures/6793 non-AE ventures) of all SBIR grant recipients. Such a prevalence rate creates 'rare-events' analysis problems in the model specification (King *et al.*, 2001). Although a rare-events model can easily accommodate the empirical model, it is computationally infeasible to calculate a shared specialization measure for 6793 non-AE ventures. To balance computational demands with empirical validity, we use a matched-pair sampling technique to compare the performance of AE and non-AE involved ventures. Specifically, we create a matched sample of 357 non-AE involved ventures using the control variables in the matching procedure. We use recently developed genetic matching algorithms that have advantages over traditional random sampling techniques, such as robustness against non-normal distributions of the control variables (for details, see (Diamond *et al.*, 2006).

4.2. Assortative Matching.

In addition to controlling for genetic matching, the AEs involved in the venture could still have unobserved factors that could affect the outcomes. Controlling for unobserved selection effects could be critical to assess the degree to which AE's self-select into SBIR related ventures. Unobserved factors such as functional specialization, years of experience, personal characteristics, or age could be some of the factors that would lead some academics into academic entrepreneurship. This suggests that, similar to matching between VC and entrepreneurs, there is a two-sided assortative matching between academics and ventures in context of academic entrepreneurship. Therefore, the matching of academics with ventures is not random but rather systematic.

We specify two assortative matching equations: (a) matching of potential academic entrepreneurs with ventures, and (ii) effects of assortative matching on venture outcomes.

Under Journal Review
Do not quote without permission

First, we start by considering relatonships between potential academic entrepreneur, i, in the NIH database over the period of observation and the pool of ventures, v. The set of links during the period of observation can be coded into an ($V \times I$) matrix, where V is the total number of ventures and I is the total number of potential academic entrepreneurs. If a potential academic entrepreneur belongs to an SBIR venture, then $p_{iv} = 1$ else it is coded zero ($p_{ivt} = 0$). As our theoretical arguments hinge on shared specialization, the assortative matching between the venture and a potential academic entrepreneur is symmetric. Therefore we eliminate the coding in the above the diagonal in the matrix. The assortative matching regression can be specified as

$$\boldsymbol{P}[p_{ivt} = 1] = \lambda[\alpha + AE_{ivt}\beta + Venture_{ivt}\delta + Industry\ Dummies\ \eta + \theta_i + \varphi_v + \varepsilon_{ivt}]$$

α represents the effect intercept, AE_{ivt} represents vector of AE characteristics [age, number of patents, AE shared specialization, AE centrality, AE knowledge combinability, and AE structural holes efficiency] five years prior to the matching event and $Venture_{ivt}$ represents the vector of venture characteristics consisting of the control variables over five years prior to matching event. θ_i is the potential AE specific fixed effect and φ_v represents venture specific fixed effects. We estimate the above specification using *felsdvreg* procedure from Stata 11. While the effects for the AE characteristics and the venture vectors were significant, the joint fixed effects were significant (F-stat potential AE and venture = 33.78, p<0.001), and potential AE fixed effects (F-stat = 18.61, p<0.001) and venture fixed effects (F-stat = 20.26, p<0.001) are significant. Therefore, there in addition to the proposed factors there is assortative matching into AE-venture pairing.

4.3. Estimation Model.

The unobserved fixed effects resulting from assortative matching, θ_i and φ_v represent the self-selection effects in line with Heckman's two step model. Therefore, we use predicted values of θ_i and φ_v to calculate the inverse-Mill's ratio to be used as a control in our specification. To estimate the effects of shared specialization and network characteristics on SBIR Phase II Award amount we use a censored Tobit model that accounts for zero dollar amounts in ventures that failed to win Phase II awards:

$$y^* = X_i\beta + \epsilon_i$$

Where $\epsilon_i \sim N(0, \sigma^2)$. y^* is a latent variable that is observed for, τ, SBIR II awards greater than zero dollars, and left-censored otherwise. The observed y is defined as:

$$y_i = \begin{cases} y^* & if\ \ y^* > \tau \\ \tau_y & if\ \ y^* \leq \tau \end{cases}$$

The second outcome variable, forward citations of patents are a nonnegative integer variable that approximates a Poisson distribution. Drawing on recommendations by Cameron *et al.* (1998), we found

Under Journal Review
Do not quote without permission

overdispersion to be a problem in the current data (LR: χ^2 (1) = 1437.852, p<0.01). Therefore, negative binomial regression is used as the estimation procedure for the forward citation dependent variable.

It is likely that ventures with patents issued in earlier periods are more likely to have higher forward citations than ventures with patents received during later periods. Clearly, the time of exposure of the patent, in addition to the underlying value of the patent, affects the number of forward citations. Similarly, it is likely that ventures that have been established for longer time periods are more likely to have more established knowledge stocks and therefore more likely to receive Phase II funding. Similarly, older ventures are more likely to have developed routines and therefore are more likely to have the organizational expertise to meet the requirements of receiving Phase II funding. To control for the effects of such time drift, we use the cross-sectional time series negative binomial regression procedure (*xtbenreg*) in Stata 11.

4.4. Empirical Findings

The means, standard deviations, and correlations for AE involved and non-AE involved ventures are presented in Table 1. All correlations are quite low and hence there does not seem to an issue of multicollinearity. Table 2 presents the censored Tobit and negative binomial results. All control variables except location legitimacy are significant. The direct effect of AE involvement is positive but insignificant in both regressions.

---------- Insert Table 1 and Table 2 about here ----------

Hypothesis 1 proposed that AE involvement when shared specialization is high will result in better venture innovation performance (Phase II award amount: β= 1.246; p<0.01; Forward patent citations (β= 1.385; p<0.01). Whether AEs with high knowledge combinability are able to contribute positively to venture innovation performance when shared specialization is high was proposed in Hypothesis 2. The results show that AEs with prior experience in broader technological domains are able to combine knowledge better, and thereby increasing the likelihood of higher Phase II Award amount (β= 1.337; p<0.01) and develop valuable patents that receive higher forward patent citations (β= 1.838; p<0.01). Finally, Hypotheses 3 and 4 proposed the positive effects of centrality and structural holes in increasing Phase II award amount (H3: β= 0.591; p<0.01; H4: β= 0.665; p<0.01) and forward patent citations (H3: β= 0.486; p<0.01; H4: β= 0.744; p<0.01). Overall, all the hypotheses are supported.

4.4. Robustness Tests

In order to further check the robustness of our results we performed a number of stress tests by using alternate estimation models, and alternative explanations for the results. We first used a censored Probit model for the SBIR Phase II Award amount. The results were consistent with the baseline results. We

Under Journal Review
Do not quote without permission

then specified a logit model for whether a venture received a Phase II award and if further patents were filed, whether the patents received forward citations. All three regressions were consistent with the baseline results. Next, we split the sample into 1996-2000 and 2001-2005 periods to assess the temporal stability of results (the year 2000 represented the start of the Internet Bubble crash). There were no significant differences in the estimates from the baseline results.

Third, following Cohen et al. (2003), we assessed the required sample size for each model with a power analysis, and used a bootstrapping method to test for coefficient stability. We report the statistical power of the individual estimations using Monte-Carlo power estimation methods in Table 2. We also report the total sample size required for power at α=0.95. Following Efron and Tibshirani (1993), we calculated power using a bootstrapping approach by drawing 1000 statistically random replications. The power for the model was 0.89. Power analysis shows that models are statistically well above the prescribed levels by Cohen et al. (2003).

Fourth, although we controlled for exposure time in patent forward citations, it is likely that increased citations are also the result of inventor network characteristics. Inventors who are more central in their networks are more likely to have more social power because they act as gatekeepers (editors, reviewers of grants, program directors, etc.) and therefore patents associated with them are likely to receive more citations (Bonacich, 1987). Therefore, there is a possibility of endogeneity between network centrality and forward citations. Following Paruchuri (2009), we use the number of collaborators before Phase I award as an instrument variable. Using a two stage least squares procedure (2SLS) we found the results to be consistent with the baseline estimations.

4.5. Post-hoc Analysis.

4.5.1. Sample Matching.

To truly assess whether AE involvement affects SBIR Phase II award outcomes or patent forward citations, one must conduct an experiment by randomly assigning ventures to two different treatments – AE involved and AE not involved. An experimental approach would randomize the distribution of venture characteristics that uniquely interact with the treatments. Thus, an experiment would provide a true test of the proposed model. But, such experiments are not practically feasible.

To address the issue of AE as a treatment effects we use propensity and genetic matching approach. One can attempt to closely replicate the experimental conditions by using matching algorithms to create a 'control' group. Sample matching is an increasingly popular method of causal inference in many fields including statistics (Rubin *et al.*, 2006), medicine (Christakis *et al.*, 2003), economics (Dehejia, 2005), political science (Hansen *et al.*, 2009), and sociology (Morgan *et al.*, 2006).

Under Journal Review
Do not quote without permission

In the current study, ventures may self-select into the treatment group, making unbiased estimates for comparison impossible. To minimize such biases, the matching algorithm must pair treatment and comparison units based on observable characteristics. Specifically, we use five matching approaches: (a) Nearest neighbor, (b) Radius matching, (c) Kernel Matching, (d) Caliper matching (δ=0.1), and (e) Genetic matching. In all the experimental conditions, the balance test was met and the results did not change under each matching algorithm.

Table 3 shows the differences in likelihood of receiving SBIR Phase II funding and increased likelihood of more patent forward citations under high shared specialization. Across all five matching algorithms we find that ventures in the AE involved treatment group displayed increased likelihoods of success for both funding and forward citation measures.

4.5.2. Counterfactual Analysis.

Continuing from the sample matching exercise, we further test for counterfactuals for ventures with high shared specialization that did not involve AEs. Drawing on the traditional econometric approach for counterfactual analysis proposed by Heckman *et al.*, (1985) and using the estimations reported in Table 2, we plot the outcomes based on counterfactuals. Because the probability of SBIR Phase II funding is a Tobit distribution, we plot a cumulative Tobit distribution of the counterfactual as reported in Figure 2(a). To assess whether there is a statistical difference, we use the Kolmogrov-Smirnov test (Δeffect = 0.097; p=0.002). We found that ventures with higher levels of shared specialization that did not involve AEs would have significantly increased their likelihood (by as high as 30%) of SBIR Phase II funding if they had AE involvement. Similarly, we assessed counterfactuals for forward citations. Figure 1(b) shows that the number of forward patent citations would have increased had ventures without AE involvement instead involved AEs. Kolmogrov-Smirnov test (Δeffect = 0.746; p=0.000) further confirms that the distributions are significantly different.

5. Discussion and Conclusion

SBIR funding is primarily geared towards advancing innovation and entrepreneurship in the basic sciences for ventures in very stages of product development. SBIR is central to promoting industry-venture collaborations. As universities are critical to advancement of basic sciences, inputs from academic entrepreneurs is fundamentally important to ventures. With limited viability of SBIR ventures beyond "hot technology" areas funded by VCs, the role of basic sciences in promoting public sector venture funding could be crucial. The parallel question in broader academic entrepreneurship literature also calls into question the importance of academic entrepreneurs.

Under Journal Review
Do not quote without permission

This paper attempts to resolve the conflicting findings in the literature on the value of academic entrepreneurs to technology ventures. The bottom line results report that AEs in high shared specialization knowledge domains are more likely to enhance biotechnology venture performance. However, a deeper consideration yields more nuanced results. Because they have specific scientific knowledge, AEs are better at creating new knowledge in local (narrow knowledge domain) searches. Those AEs with prior experience in diverse technological domains are more likely to enhance venture innovation performance. Hence, those AEs with high knowledge combinability capabilities can increase the effectiveness of their local searches by accessing diverse knowledge stores held in their networks.

Our findings are also important to the work on knowledge networks. Earlier studies on social networks have not examined their dynamics from an inter-network and intra-individual perspective. By considering knowledge combinability and network structures in the context of an actor's knowledge domain, we find that the value of a network position that is central and rich in structure holes is further increased by an actor's ability (through high knowledge combinability) to overcome technological distances.

We find that centrality enhances variations in solutions generated and increases search efficiency in distant searches. Similarly, by positioning themselves in structural holes AEs can more effectively sift diverse information from distant searches. Compared to previous work on AEs, we proposed that the contingency related to narrow or broad technological domains is central to AE effectiveness. In addition to the role of human capital proposed by Toole and Czarnitzki (2009), network capital and the ability to combine diverse knowledge seem to be as important to an AE's effectiveness when operating in broad technology domains.

Prior research exploring the importance of knowledge flows and knowledge diversity in the context of centrality and structural holes (Ahuja, 2000) has suggested close examinations of the characteristics of the node to better explain knowledge accessibility and transfer. Recent work by Paruchuri (2010) and others has focused on role of individual positions in their knowledge networks. Our findings add to this line of inquiry by suggesting that the value of a network position is contingent on the knowledge domain in which an actor finds herself.

From a policy perspective, the findings suggest that SBIR award criteria could focus more closely on the shared specialization of the venture and the shared specialization of the academic entrepreneur and the venture. Ventures with shared specialization, or innovations based on path dependent knowledge building are more likely to develop viable innovations. Faced with limited resources and capabilities, shared specialization provides "innovation niche" that is more defensible and viable for a venture. Ventures exploiting innovations with lower levels of shared specialization or high technological diversity seem less likely to develop innovations. More importantly, on the partnership side, dyadic shared

Under Journal Review
Do not quote without permission

specialization or shared specialization between ventures and academic entrepreneurs could be critical to joint commercialization efforts. Mismatches in the degree of specialization could significantly limit knowledge recombination efforts.

This work is not without the standard limitations. Academic entrepreneurship is an emerging research area. With complex processes such as knowledge development in ventures, we know that a number of exogenous and endogenous factors can also affect innovation success. Therefore, we cannot claim that causality is implied by the results, although our counterfactual analysis suggests our arguments are sound. Second, networks are composed of complex knowledge exchange mechanisms and the current analysis only looks at the structural aspects of network positions. It is likely that individuals with similar network positions may not enjoy similar gains because of individual differences such as emotional intelligence. The homogenous assumption of network actors is a general problem in the theoretical literature. It places a limit on how far we can take our conclusions since the capabilities of AEs are at the heart of this paper. Future research may focus on such complex phenomena as the knowledge exchange dynamics in ventures. Finally, the findings are not generalizable to contexts outside the US or even of the technological domain (biotechnology). This is because the research process, patent laws, the patenting process, and the granting process are different. Despite some of these limitations, we believe that our findings offer new insights on the reasons why the success of AEs may be contingent on their network positions and knowledge combination capabilities.

Under Journal Review
Do not quote without permission

References

Agrawal A. 2006. Engaging the inventor: Exploring licensing strategies for university inventions and the role of latent knowledge. *Strategic Management Journal* **27**(1): 63-79

Ahuja G. 2000. Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*: 425-455

Audretsch DB. 2003. Standing on the shoulders of midgets: The US Small Business Innovation Research program (SBIR). *Small Business Economics* **20**(2): 129-135

Audretsch DB, Link AN, Scott JT. 2002a. Public/private technology partnerships: evaluating SBIR-supported research* 1. *Research Policy* **31**(1): 145-158

Audretsch DB, Weigand J, Weigand C. 2002b. The impact of the SBIR on creating entrepreneurial behavior. *Economic Development Quarterly* **16**(1): 32

Berger PL, Luckmann T. 1967. *The social construction of reality*. Doubleday New York

Bessette RW. 2003. Measuring the economic impact of university-based research. *The Journal of Technology Transfer* **28**(3): 355-361

Bonacich P. 1987. Power and centrality: A family of measures. *American Journal of Sociology* **92**(5): 1170-1182

Brass DJ, Burkhardt ME. 1993. Potential power and power use: An investigation of structure and behavior. *Academy of Management Journal* **36**(3): 441-470

Burt RS. 1995. *Structural holes: The social structure of competition*. Harvard Univ Pr

Christakis NA, Iwashyna TJ. 2003. The health impact of health care on families: a matched cohort study of hospice use by decedents and mortality outcomes in surviving, widowed spouses. *Social Science & Medicine* **57**(3): 465-475

Cockburn I, Henderson R. 1998. Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *The Journal of Industrial Economics* **46**(2): 157-182

Cohen J, Cohen P, West SG, Aiken LS. 2003. *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences.* Lawrence Erlbaum Associates, Inc: Mahwah

Dehejia R. 2005. Practical propensity score matching: a reply to Smith and Todd. *Journal of Econometrics* **125**(1-2): 355-364

Diamond A, Sekhon JS. 2006. Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Working paper*

Under Journal Review
Do not quote without permission

Dushnitsky G, Lenox MJ. 2005. When do firms undertake R&D by investing in new ventures? *Strategic Management Journal* **26**(10): 947-965

Efron B, Tibshirani R. 1993. *An introduction to the bootstrap*. Chapman & Hall/CRC

Feldman MP. 2000. Where science comes to life: university bioscience, commercial spin-offs, and regional economic development. *Journal of Comparative Policy Analysis* **2**(3): 345-361

Fleming L, Sorenson O. 2004. Science as a map in technological search. *Strategic Management Journal* **25**(8-9): 909-928

Gans JS, Stern S. 2000. When does funding research by smaller firms bear fruit?: Evidence from the SBIR program. National Bureau of Economic Research Cambridge, Mass., USA

Gilsing V, Nooteboom B, Vanhaverbeke W, Duysters G, Van Den Oord A. 2008. Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density. *Research Policy* **37**(10): 1717-1731

Gittelman M, Kogut B. 2003. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. *Management Science* **49**(4): 366-382

Hansen BB, Bowers J. 2009. Attributing Effects to a Cluster-Randomized Get-Out-the-Vote Campaign. *Journal of the American Statistical Association* **104**(487): 873-885

Heckman JJ, Robb Jr R. 1985. Alternative methods for evaluating the impact of interventions:: An overview. *Journal of Econometrics* **30**(1-2): 239-267

Ibarra H. 1993. Personal networks of women and minorities in management: A conceptual framework. *Academy of Management review* **18**(1): 56-87

Kaplan SN, Sensoy BA, Strömberg P. 2009. Should investors bet on the jockey or the horse? Evidence from the evolution of firms from early business plans to public companies. *The Journal of Finance* **64**(1): 75-115

Kauffman SA, Weinberger ED. 1989. The NK model of rugged fitness landscapes and its application to maturation of the immune response*. *Journal of Theoretical Biology* **141**(2): 211-245

Kogut B. 2000. The network as knowledge: generative rules and the emergence of structure. *Strategic Management Journal* **21**(3): 405-425

Laursen K, Salter A. 2006. Open for innovation: the role of openness in explaining innovation performance among UK manufacturing firms. *Strategic Management Journal* **27**(2): 131-150

Leblebici H, Salancik GR. 1981. Effects of environmental uncertainty on information and decision processes in banks. *Administrative Science Quarterly*: 578-596

Under Journal Review
Do not quote without permission

Lerner J. 1999. The government as venture capitalist: The long-run impact of the SBIR program. *Journal of business*: 285-318

Morgan SL, Harding DJ. 2006. Matching estimators of causal effects. *Sociological Methods & Research* **35**(1): 3-60

Nooteboom B. 2000. Learning by interaction: absorptive capacity, cognitive distance and governance. *Journal of Management and Governance* **4**(1): 69-92

Paruchuri S. 2010. Intraorganizational Networks, Interorganizational Networks, and the Impact of Central Inventors: A Longitudinal Study of Pharmaceutical Firms. *Organization Science* **21**(1): 63-80

Rubin DB, Stuart EA. 2006. Affinely invariant matching methods with discriminant mixtures of proportional ellipsoidally symmetric distributions. *The Annals of Statistics* **34**(4): 1814-1826

Shane S. 2000. Prior knowledge and the discovery of entrepreneurial opportunities. *Organization Science*: 448-469

Shane SA. 2004. *Academic entrepreneurship: University spinoffs and wealth creation*. Edward Elgar Publishing

Stern S. 2004. Do scientists pay to be scientists? *Management Science* **50**(6): 835-853

Stinchcombe AL. 1965. Organizations and social structure. *Handbook of organizations*: 142-193

Stuart TE, Podolny JM. 1996. Local search and the evolution of technological capabilities. *Strategic Management Journal* **17**(S1): 21-38

Toole A, Czarnitzki D. 2007a. Biomedical academic entrepreneurship through the SBIR program. *Journal of Economic Behavior & Organization* **63**(4): 716-738

Toole A, Czarnitzki D. 2009. Exploring the relationship between scientist human capital and firm performance: the case of biomedical academic entrepreneurs in the SBIR program. *Management Science* **55**(1): 101-114

Toole AA, Czarnitzki D. 2007b. Biomedical academic entrepreneurship through the SBIR program. *Journal of Economic Behavior & Organization* **63**(4): 716-738

Von Wartburg I, Teichert T, Rost K. 2005. Inventive progress measured by multi-stage patent citation analysis. *Research Policy* **34**(10): 1591-1607

Yayavaram S, Ahuja G. 2008. Decomposability in knowledge structures and its impact on the usefulness of inventions and knowledge-base malleability. *Administrative Science Quarterly* **53**(2): 333-362

Under Journal Review
Do not quote without permission

Zucker L, Darby M. 1996. Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. *Proceedings of the National Academy of Sciences of the United States of America* **93**(23): 12709-12716

Zucker L, Darby M, Armstrong J. 2002. Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology. *Management Science* **48**(1): 138-153

Zucker L, Darby M, Brewer M. 1998. Intellectual human capital and the birth of US biotechnology enterprises. *American Economic Review* **88**(1): 290-306

Under Journal Review
Do not quote without permission

TABLE 1

Mean, SDs, zero-order correlation Tables

	Mean	SD	1	2	3	4	5	6
1. AE	0.500[10]	0.000	1					
2. $AE_{Centrality}$	0.271	0.209	0.095	1				
3. $AE_{Structural\ holes\text{-}constraint}$	0.480	0.468	0.081	0.508	1			
4. $AE_{Knowledge\ Combinability}$	86.493	39.352	0.093	0.217	0.311	1		
5. Phase II Award Amount	0.061	0.197	0.082	0.081	0.021	0.124	1	
6. Forward Citations	29.37	29.40	0.103	0.069	0.019	0.113	0.241	1

		AE-Mean	AE-SD	Non-AE Mean	Non-AE SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	Shared Specialization Measure	0.398	0.155	0.201	0.442	1	0.508	0.526	-.101	-.117	.056	.127	.111	.022	.031	.027	.132	.106	.101
2.	Reachability out-degree measures	0.503	0.262	0.314	0.369	0.419	1	0.357	-0.097	-0.223	0.073	0.127	0.241	0.092	0.032	0.035	0.225	0.242	0.149
3.	Bibliographic Coupling	0.293	0.178	0.157	0.288	0.442	0.338	1	-0.134	-0.087	0.068	0.155	0.096	0.044	0.041	0.016	0.132	0.133	0.155
4.	External Legitimacy	0.384	0.407	0.472	0.276	-.187	-0.253	-0.136	1	.090	.054	.064	.108	.094	.034	.028	.062	.114	.192
5.	Patent Stock (t-1)	1.153	5.824	0.721	3.461	-.128	-0.333	-0.077	.084	1	.015	.089	.086	.092	.058	.014	.056	.068	.138
6.	Location Legitimacy	15.62	17.39	16.20	18.71	.075	0.240	0.045	.046	.013	1	.012	.032	.067	.010	.012	.038	.017	.043
7.	Firm Size	8.46	2.73	11.73	4.28	.088	0.352	0.063	.070	.075	.014	1	-.054	.110	.026	.016	.071	.096	.077
8.	Competitive intensity	16.34	11.50	28.74	17.02	.147	0.353	0.112	.117	.099	.024	-.064	1	-.022	.014	.014	.052	.072	-.066
9.	VC Backed (t-1)	0.156	0.404	0.077	0.281	.041	0.164	0.021	.086	.083	.045	.097	-.012	1	.029	.017	.160	.131	.125
10.	Minority-owned	0.073	0.297	0.099	0.346	.089	0.320	0.043	.022	.044	.017	.023	.018	.012	1	.038	.065	.042	.027
11.	Female-owned	0.103	0.355	0.109	0.286	.028	0.106	0.013	.030	.026	.012	.014	.011	.018	.021	1	.082	.094	.115
12.	Phase 1 $(t-1)	6.417	3.057	8.562	2.597	.159	0.509	0.076	.074	.062	.025	.065	.043	.137	.082	.105	1	.115	.232
13.	Phase II Award Amount	8.442	2.844	7.447	2.503	.199	0.216	0.088	.093	.041	.014	.089	.059	.152	.099	.082	.182	1	.104
14.	Forward Citations	29.37	29.40	32.81	27.44	.175	0.385	0.105	.174	.115	.031	.072	.042	.166	.063	.124	.193	.242	1

Lower diagonal represents correlations for AE-linked firms, and the upper diagonal represents correlations for non-AE linked firms.
All correlations above 0.08 are significant at $p<0.05$ or lower (two-tailed test)

[10] Due to Genetic matching without replication the ratio of AE-linked and non-AE linked firms is 50-50.

Under Journal Review
Do not quote without permission

TABLE 2

Probit and Tobit Regression for SBIR Phase II award dummy and SBIR Award Amount[11]

Variable	Tobit Model Dependent Variable: Phase II Award Amount [2,12]	Statistical Power (alpha=0.05)	Negative Binomial Regression Dependent Variable: Forward Citations[13]	Statistical Power (alpha=0.05)
Direct Effects				
AE	1.546 (0.353)***	0.92	1.322 (0.188)***	0.93
Shared Specialization	1.705 (0.046)***	0.86	1.576 (0. 259)***	0.86
$AE_{Centrality}$	0.236 (0.087)***	0.93	0.144 (0.057)***	0.89
$AE_{Structural\ holes\text{-}efficiency}$	0.368 (0.117)***	0.87	0.436 (0.095)***	0.92
$AE_{Knowledge\ Combinability}$	0.304 (0.079)***	0.91	0.315 (0.113)***	0.85
Moderation Effects				
Shared Specialization × AE [H1]	1.246 (0.568)***	0.91	1.385 (0.226)***	0.88
Shared Specialization × $AE_{Knowledge\ Combinability}$ [H2]	1.337 (0.419)***	0.93	1.838 (0.385)***	0.92
Shared Specialization × $AE_{Centrality}$ [H3]	0.591 (0.197)***	0.83	0.665 (0.358)*	0.90
Shared Specialization × $AE_{Structural\ holes\text{-}efficiency}$ [H4]	0.486 (0.185)***	0.91	0.744 (0.236)***	0.91
Controls				
External Legitimacy	0.457 (0.217)**	0.85	0.361 (0.151)**	0.84
Patent Stock (t-1)	0.241 (0.077)***	0.91	0.153 (0.056)***	0.89
Location Legitimacy	1.357 (0.867)	0.89	1.474 (0.984)	0.84
Firm Size	9.592 (4.638)**	0.84	5.327 (6.438)	0.81
Competitive intensity	33.684 (9.374)***	0.81	8.468 (4.762)*	0.93
VC Backed (t-1)	0.224 (0.087)*	0.92	0.157 (0.060)*	0.85
Minority-owned	0.026 (0.057)	0.93	0.078 (0.061)	0.87
Female-owned	0.198 (0.281)	0.87	0.249 (0.121)*	0.86
Phase 1 $(t-1)	4.438 (1.038)***	0.86	2.568(1.176)**	0.90
Reachability out-degree measures	1.692 (0.244)***	0.93	1.275 (0.366)**	0.90
Bibliographic Coupling	1.487 (0.576)***	0.92	1.153 (0.279)**	0.90
Inverse-Mills ratio -- θ_i	0.132 (0.051)**	0.86	0.227 (0.109)*	0.87
Inverse-Mills ratio -- φ_v	0.244 (0.119)**	0.83	0.281 (0.132)**	0.85
10 Time Dummies (1996-2005)	Yes		Yes	
17 technological class dummies	Yes		Yes	
Intercept	1.632 (0.421)***		1.217 (0.424)***	

[11] *Notes.* Heteroskedastic Probit and Tobit models; heteroskedasticity is modeled groupwise multiplicatively; heteroskedasticity term includes 10 time dummies and 17 field dummies. Firm-clustered standard errors are in parentheses. ***, **, and * indicate a significance level of 1%, 5%, and 10%, respectively. bNegative binomial regressions. Standard errors are calculated heteroskedastic-consistently. Equidispersion is clearly rejected: LR(χ^2 (1) = 1437.852

[12] *xtbenreg* in Stata to control for venture exposure time – venture age [years since venture establishment; derived from Corptech directory].

[13] *xtbenreg* in Stata to control for venture patent exposure time – average age of patents in venture portfolio prior to receiving SBIR award.

Under Journal Review
Do not quote without permission

Joint Significance of AE and AE network characteristics [$\chi^2(5)$]	19.322		22.148	
Joint significance of AE, AE network characteristics, and moderation effects [$\chi^2(9)$]	25.394		36.208	
Joint Significance of technology classes [$\chi^2(17)$]	62.486		39.044	
Joint Significance of time dummies [$\chi^2(10)$]	36.122		48.669	
Number of Observations	714		714	
McFadden R-Square	0.562		0.464	
Log-likelihood	-5,322.404		-6,986.536	

* *p<0.05*
** *p<0.01*
*** *p>0.001*

Under Journal Review
Do not quote without permission

TABLE 3

Propensity Score Matching and Genetic Matching Algorithms

	Differences in Phase II Award ($AE/non-AE$): shared specialization				Differences in Forward citations ($AE/non-AE$): shared specialization			
	Δ	s.e.[14]	t-value	p-value	Δ	s.e.	t-value	p-value
Propensity Score Matching[15,16]								
Nearest Neighbor Matching	0.156	0.077	2.026	0.021	0.181	0.045	4.022	0.000
Radius Matching	0.243	0.030	8.100	0.000	0.287	0.049	5.857	0.000
Kernel Matching	0.296	0.079	3.747	0.000	0.226	0.048	4.708	0.000
Caliper matching (δ=.1)	0.220	0.076	2.895	0.002	0.168	0.038	4.421	0.000
Genetic Matching Algorithm [3, 17]								
GenMatch	0.168	0.065	2.585	0.005	0.278	0.090	3.089	0.001

[14] All bootstrap standard errors are based on 1000 replications
[15] *psmatch2* in Stata 11
[16] Balancing hypothesis [we.e. distribution of control (0=non-AE involvement) and treatment (1=AE involvement) groups are identical for the matching covariates] was supported for *psmatch2* and *Matching* algorithms.
[17] R package *Matching* for genetic algorithm matching

Under Journal Review
Do not quote without permission

Figure 1
Counterfactual Analysis.

Figure 1(a): Counterfactual plots for non-AE involving ventures with high shared specialization – SBIR Phase II funding.

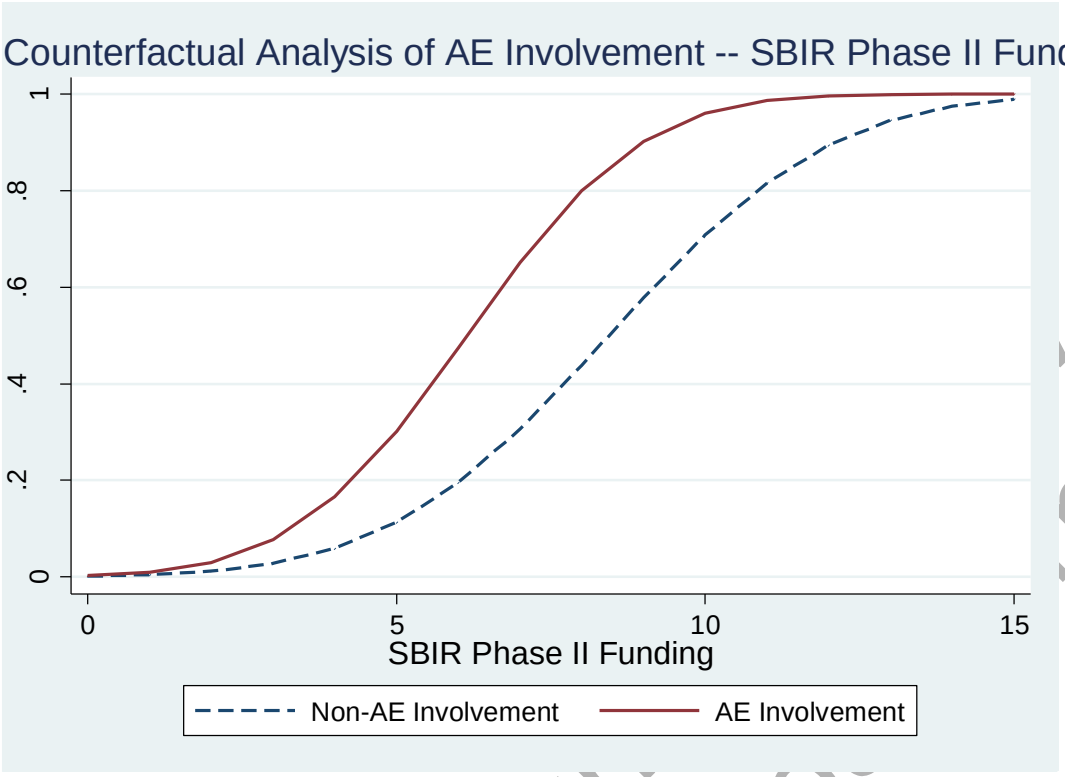
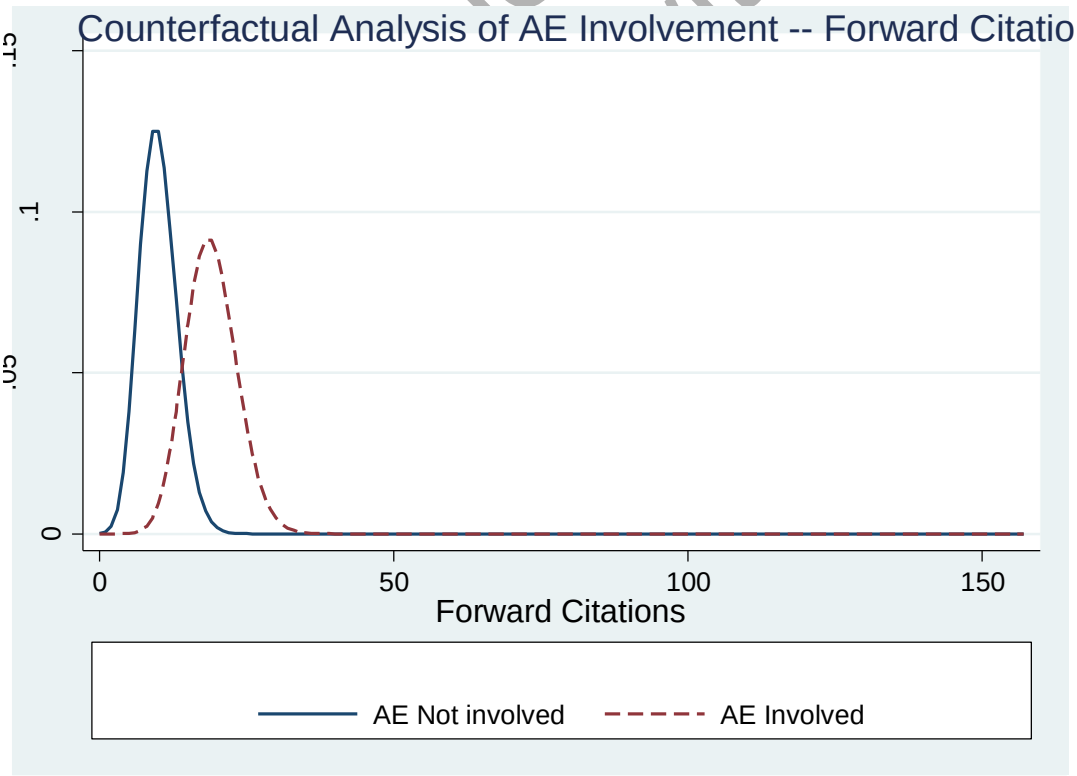


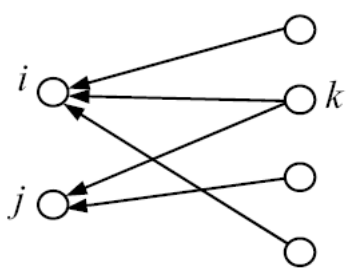
Figure 1(b): Counterfactual plots for non-AE involving ventures with high shared specialization – Forward Citations.



Appendix 1

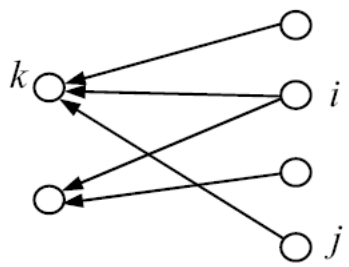
Figure A.1.1. Bibliographical Coupling versus Co-citation

Co-citation
If patent *i* and patent *j*, both are cited by patent *k*, then the degree of co-citation between patent *i* and *j* is given as, $C_{ij} = \sum_{k=1}^{n} L_{ki} L_{kj}$, where L is the citation matrix, and $L_{ki} = 1$, if patent *k* cites patent *i*, else zero (same for L_{jk})



Both patent *i* and patent *j* are co-cited by patent *k*.

Bibliographic Coupling
If patent *i* and patent *j*, both cite patent *k*, then the degree of bibliographic coupling between patent *i* and *j* is given as, $B_{ij} = \sum_{k=1}^{n} L_{ik} L_{jk}$, where L is the citation matrix, and $L_{ik} = 1$, if patent *i* cites patent *k*, else zero (same for L_{jk})

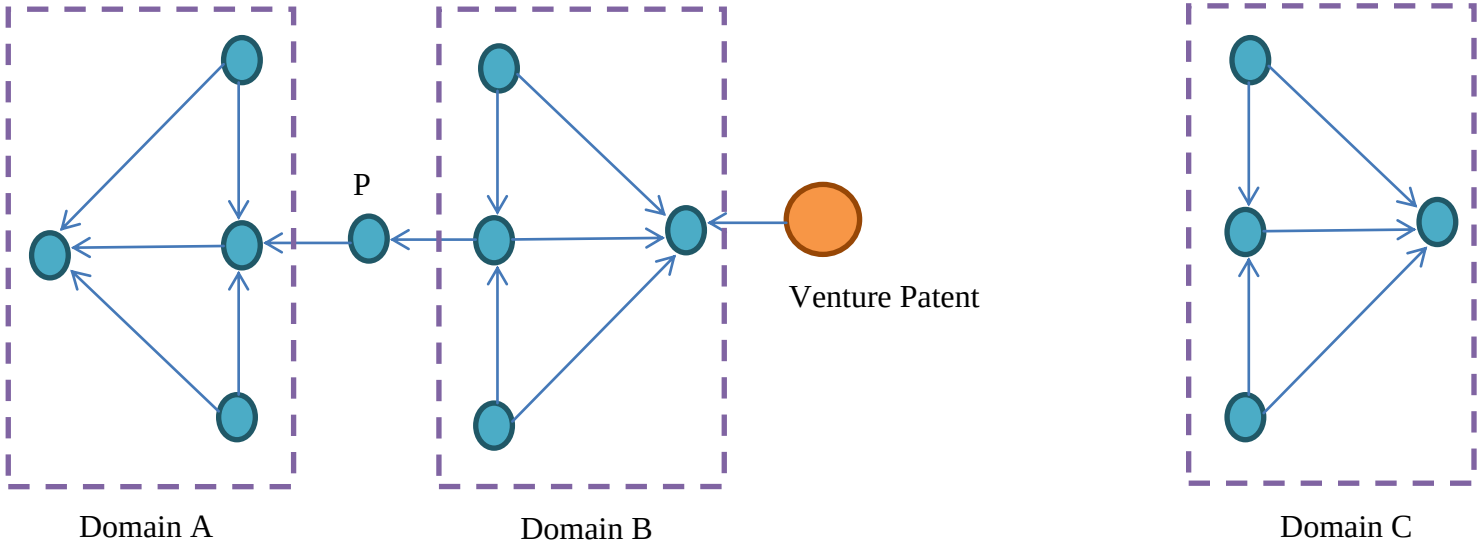


Both patent *i* and patent *j* cite patent *k*.

Under Journal Review
Do not quote without permission

Figure A.1.2. Reachability in Networks

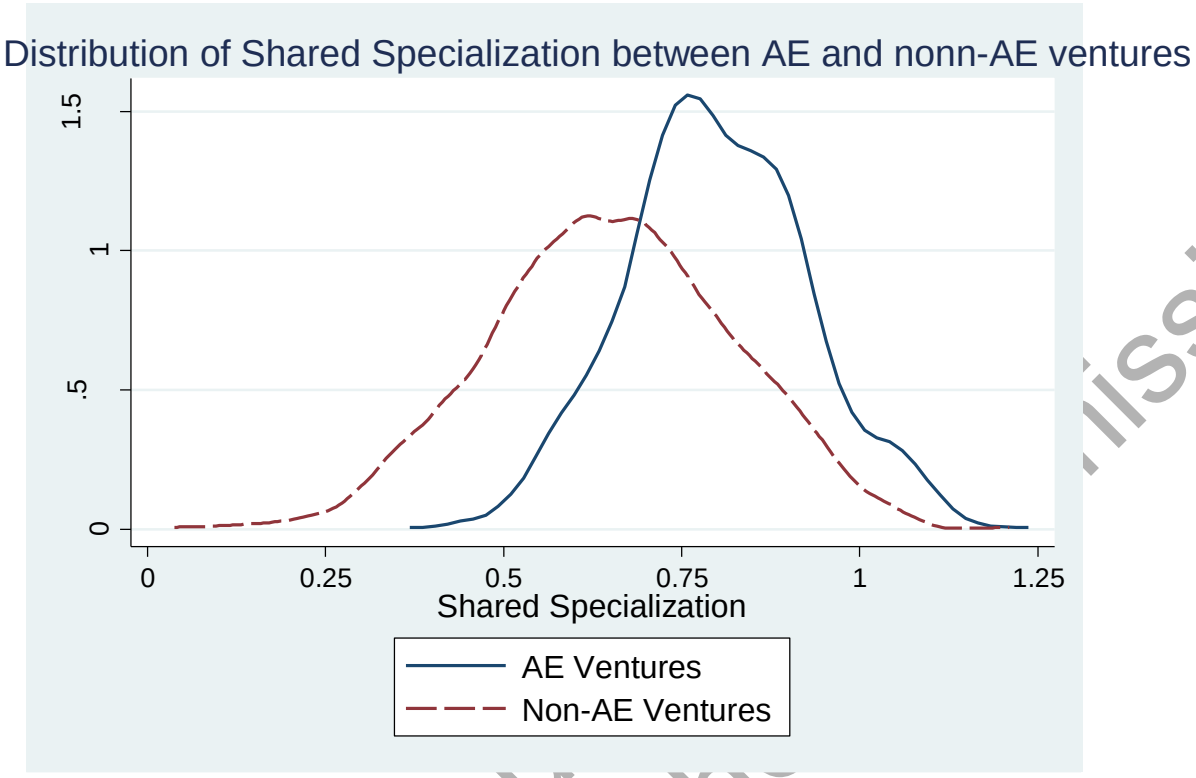
Out-degree Reachability is the degree to which a ventures patent network is connected directly or indirectly to other patents of any length. If out-degree reachability is limited then there is disruption in the flow of knowledge and it is likely that the venture patent portfolio is a part of a sub-network of knowledge domain.



Venture is connected to Domains A through patent P, whereas the venture has zero out-degree reachability to Domain C.

Under review
Do not quote

Figure A.1.3. Shared specialization between AE involved and non-AE involved ventures



Mann-Whitney test: z=2.137 [AE ventures minus non-AE ventures] p=0.016 (two-tailed)

Under Journal Submission
Do not quote without