Network-Mediated Knowledge Spillovers in ICT/Information Security

March 31, 2018

Lee Branstetter
Neil Gandal
Nadav Kunievsky

Abstract. A large and growing literature has used patent and patent citation data to measure knowledge spillovers across inventions and organizations, but this literature has not explicitly considered the collaboration networks formed by inventors as a mechanism for shaping and transmitting these knowledge flows. This paper examines the incidence and nature of knowledge flows mediated by the collaboration networks of inventors active in the information and communication technology (ICT) and information security sectors.

Using data from U.S. PTO patent grants from ICT patent classes that include information security patents, we find that the quality of Israeli inventions in this area is systematically linked to the structure of the collaborative network generated by Israeli inventors in this sector. This suggests that there are knowledge spillovers from the Israeli network. This research highlights the importance of direct interaction among inventors as a conduit for flows of frontier scientific knowledge.

Branstetter, Gandal, and Kunievsky thank the Maurice Falk Institute for Economic Research in Israel and Start-Up Nation Central for financial support of this research. We thank Britta Glennon for excellent research assistance, and are grateful to Eugene Kandel and seminar participants at Collegio Carlo Alberto, Hebrew University, UC-Berkeley and Stanford for helpful comments and suggestions.

© 2018 by Lee Branstetter, Neil Gandal, and Nadav Kunievsky. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including the © notice, is given to the source.
1. **Introduction**

Knowledge spillovers lie at the heart of modern theories of endogenous growth (Romer, 1986, 1990; Acemoglu, 2009), international trade (Grossman and Helpman, 1991; Branstetter and Saggi, 2011); international investment (Keller and Yeaple, 2013), and economic development (Jones, 2014). The late Zvi Griliches and several generations of his students, including Adam Jaffe and Manuel Trajtenberg (2002), introduced a series of econometric techniques for empirically measuring the strength of these spillovers across time and space, using patents and patent citations. A large and growing literature has deployed these techniques across a wide range of technological domains, organizational categories, and countries, strongly affirming the existence and importance of knowledge spillovers.¹

Despite this extensive literature, the exact mechanisms through which knowledge spillovers are propagated, their relative importance in mediating these knowledge flows - and the effects of these spillovers on the quality of the end products - remain imperfectly understood. Some early research (Griliches, 1979, 1992; Keller, 1998) presumed that at least some spillovers might flow through contact in the marketplace with products or services embodying new technology. Other firms might reverse-engineer and build on this technology without ever forging any direct contact between their R&D engineers and those of the firm that created the original product. While this kind of spillover is certainly possible, in modern technology-intensive industries, spillovers are also likely to occur through more direct interaction between individuals who work together and exchange ideas and information.

High-tech R&D is typically done by teams. Working in teams necessarily involves exchanging ideas and sharing information. Participants of such research teams carry this knowledge to other teams and other projects in which they are involved or become involved, and knowledge can continue to flow between former collaborators even after they move across regions or to different firms and cease direct collaboration (Almeida et al., 2001; Agrawal et al., 2006). The networks traced out by collaborations can become a key mechanism through which knowledge flows. Interestingly, though a great deal of the research has focused on measuring knowledge spillovers, which is not the focus of the current paper.

¹ The empirical literature on knowledge spillovers is quite extensive, and we lack the space to review it fully. Scherer (1982), Jaffe (1986), Bernstein and Nadiri (1988), and Irwin and Klenow (1994) authored influential early studies, and Griliches (1992) provided a survey of early empirical work. Keller (2004) provides a review of the empirical literature focused on international knowledge spillovers, which is not the focus of the current paper.
spillovers in patents, over time and space, to the best of our knowledge, no previous research has tried to link knowledge spillovers in the networks formed by inventors' joint work to the quality of patents.

In this paper, we apply a model developed by Fershtman and Gandal (FG 2011) (and applied to Open Source Software) to examine the existence and importance of collaborator network-mediated knowledge spillovers in the ICT/information security industry. This is an industry in several nations outside the United States, have emerged as important centers of innovation.

Using data from U.S. PTO patent grants in information security, we find that the quality of ICT/information security inventions is systematically linked to the structure of the collaborative network in the case of Israel. This suggests that there are knowledge spillovers in the Israeli network, which improve the quality of patents, as measured by the number of citations. This research highlights the importance of direct interaction among inventors as a conduit for flows of frontier scientific knowledge.

1.1 Literature Review

Our paper is related to two strands of literature. The first strand, pioneered by Trajtenberg (1990), uses patent citations as measures of the quality of innovations and as measures of knowledge spillovers across inventions. More important inventions tend to be cited more frequently by subsequent patents, in the same way that important and influential papers receive more citations from later scholarship. Empirical techniques initially developed by Jaffe, Trajtenberg, and Henderson (1993) and reviewed in Jaffe and Trajtenberg (2002) use patent citations to measure knowledge spillovers across time and space. As this literature evolved, a growing number of papers sought to directly measure social, contractual, or institutional connections between inventors that might mediate knowledge spillovers between them. Branstetter (2001, 2006), Singh (2008), Berry (2012), and Alcacer and Zhao (2012), among others, built on the techniques of Jaffe, Trajtenberg, and Henderson, and used them to measure the degree to which multinationals can enhance flows of knowledge spillovers across national boundaries by creating R&D facilities abroad. Gomes-Casseres, Hagedoorn and Jaffe (2006) and Branstetter and Sakakibara (2002) have used patent and citation data to measure the impact of formal interfirm research collaboration on knowledge spillovers. Almeida et al. (2001) and Agrawal, Cockburn, and McHale (2006), among many others, have sought to measure the
impact of the movement of specific individual inventors across organizational boundaries on knowledge spillovers between them. Interestingly, however, virtually no previous studies in the economics literature have examined the impact of inventors' collaboration network traced out by coinventions (that is, inventors appearing together previously on the same patent document) on knowledge flows and invention quality.2

This omission in the innovation literature is striking given the significant attention placed on collaboration networks in other, closely related social science literatures. Recent studies have examined the relationship between network structure and behavior (e.g., Ballester, Calvó-Armengol, & Zenou, 2006; Calvo-Armengol & Jackson, 2004; Goyal, van der Leij and Moraga-Gonzalez, 2006; Jackson & Yariv, 2007; Karlan, Mobius, Rosenblat, & Szeidl, 2009) and the relationship between network structure and performance (Ahuja, 2000; Calvó-Armengol, Patacchini, & Zenou, 2009, Fershtman and Gandal, 2011, and Gandal and Stettner, 2016). This paper seeks to fill a gap in the literature by assessing the degree to which collaboration networks, as traced out by pre-existing instances of “coinvention” by engineers named in patent documents, shape the pattern of knowledge spillovers and influence patent quality.

1.2 Our Analysis and Results

In this paper, we use data on the inventors that appear in patent documents to trace out and construct a two-mode network: (I) a Patent network and (II) an Inventor network. In the case of the patent network, the nodes are the patents and two patents are linked if there are inventors who work in both. In the case of the inventor network, the nodes of this network are the inventors themselves. There is a link between two inventors if they jointly hold a patent. (In section 3 below we provide a simple example to distinguish these two networks.)

We examine the patent network and the inventor (collaboration) network of inventors creating technologies in the domain of information security, broadly defined. Our broad definition includes all patents in ICT patent classes that the USPTO defines as information security related classes; these are listed in detail in Appendix A, and discussed later in the paper. For each

---

2 Breschi and Lissoni (2009) provide an exception. Their question and approach differs ours. They are primarily interested in distinguishing knowledge flows that are due to (1) local proximity versus those due to (2) inventors who move from firm to firm locally. While they build a co-invention network, they do not formally use the properties of the network in the analysis, and do not link structural characteristic of the network to the quality of patents.
patent, we calculate its proximity to other patents in the network, where the links are through inventors. We then calculate the centrality of these patents within patent network, in a manner defined below. Similarly, we calculate the centrality of inventors within the inventor network.

We then regress patent invention quality, measured by the total number of forward citations, on network centrality measures within the patent network at the time when the patent application was submitted. We control for other characteristics of the patent. We find that in the case of Israel, the network centrality measures are significantly associated with the variation in patent quality. In the context of the FG (2011) model, this result provides evidence of both direct and indirect knowledge spillovers.

We use instances of “coinvention” – the same inventors appearing together in a patent document – to trace out the networks through which knowledge spillovers will be presumed to flow. Of course, this definition necessarily omits instances of collaboration or communication that are not reflected in the “paper trail” left by coinvention. While acknowledging this point, we argue that unmeasured communication and interaction is likely to be highly correlated in space and time with the coinvention episodes that we do observe in the patent data record.

1.3 Israel's Emergence as a Global Center of Innovation in ICT/Information Security

Our primary focus is on Israel, which is recognized as one of the most innovative countries in the world. (We will examine other countries as well in future drafts.) A key initial element in this is Israel's innovative environment. Widely cited indices of national innovative capacity, such as the Bloomberg Index of Innovation or the Global Competitiveness Index compiled by the World Economic Forum, regularly rank Israel among the world’s top 5 innovating countries, despite its small size.\(^3\) Reflecting this technological strength, the country has become a major global center for high-tech entrepreneurship. Excluding the U.S., only China has more firms listed on the NASDAQ stock exchange.\(^4\) Leading players in the global IT sector, such as Intel, IBM, Google, Motorola, Apple, Microsoft, and many others have set up


research centers in Israel, hoping to harvest local talent and knowledge. Israeli companies today play a key role in shaping the global IT industry - from chips to the end user applications. Israeli firms occupy an especially prominent role in information security, which is one of the largest and fastest growing sub-sectors of ICT.

Popular explanations of Israel’s technological ascendance characterize Israel’s size as a strength, asserting that the small nation is characterized by tightly connected networks, through which knowledge spillovers can easily flow. Elite Israel Defense Force (IDF) units, such as the well-known Unit 8200, are believed to play an important role in seeding successful startups in Israel by creating a connected network of programmers.\(^5\) Unit 8200, and similar units, effectively nudge a fraction of their most gifted alumni into high-tech entrepreneurship in ICT and related domains. Once they leave the military, 8200 veterans use the network of 8200 veterans to found start-ups and develop technologies based in part on their experience and connections in the military.\(^6\) The theme of knowledge spillovers from connected networks of former members of the military intelligence corps runs through the book *Start-Up Nation* (Senor and Singer 2009) and other sources, but no rigorous work has been conducted on this issue.

In this paper, we do not address the role of particular military units in fostering Israeli networks of information technology developers. However, we undertake what is, to the best of our knowledge, the first empirical effort to measure these networks, as they are traced out in patent data, and ascertain the degree to which network density affects the quality of Israeli invention.

To capture information security inventions, we include all patents granted within a broad range of ICT patent classes that have been identified by the USPTO as containing information security patents. These classes are reasonably broad, and contain within them patents that are not strictly information security inventions, per se. It was important for us to include all of

\(^5\) Unit 8200, a military intelligence unit focusing on signal intelligence and code decryption, is the largest unit in the Israel Defense Forces, comprising several thousand soldiers. It is comparable in its function to the United States’ National Security Agency. See Idan Tendler, “From the Israeli Army Unit 8200 to Silicon Valley,” 23 March 2015, available at https://techcrunch.com/2015/03/20/from-the-8200-to-silicon-valley/

\(^6\) “70 percent of successful Israeli startups are led by 8200 graduates,” says NBIC Director Fadi Swidan,” from “High-tech elites to nurture Arab-Israeli startups,” 17.4.2016, available at http://www.israel21c.org/high-tech-elites-to-nurture-arab-israeli-startups/
information security classes as defined by the USPTO. Additionally, very narrowly defined fields have limited numbers of patents and make econometric work infeasible.

Finally, Israel is very different from the other countries because a large proportion of its patents in the ICT/Information Security sector (47 percent) are assigned to US firms. No other country in the database has more than 17 percent US assignees, and most of the countries have less that 5 percent US assignees. Since many patents with Israeli inventors have US assignees, we separately examine Israeli patents with US Assignees and Israeli patents with Israeli assignees separately. We find similar levels of spillovers between these two groups.

2. Theoretical Foundations for Network-Mediated Knowledge Spillovers

Network-mediated knowledge spillovers can be either direct or indirect. In the case of network-mediated spillovers between patented inventions, direct spillovers occur when two patented inventions have a common inventor who transfers knowledge from one patent to another. That is, an inventor takes the knowledge that he/she acquired while working on a previously patented invention and implements it in another invention. However, knowledge may also flow between invention teams even if they are not directly connected by a common inventor. The indirect route occurs whenever an inventor learns something from participating in one invention, takes the knowledge to a second invention and "shares" it with another inventor on that invention team, who, in turn, uses it when she works on a third invention. In such a scenario, knowledge flows from the first patent to the third patent, even though they do not have any inventors in common. Clearly, such indirect spillovers may be subject to decay depending on the distance (the number of the indirect links) between the patents.

Fershtman and Gandal (FG 2011) show theoretically that whenever there are direct spillovers, there should be a positive correlation between project success and the degree of the project (the number of projects with which the focal project has a common developer). When there are both direct and indirect project spillovers, there should be a positive correlation between project success and project closeness centrality, which is defined as the inverse of the sum of all distances between the project and all other projects. Closeness centrality thus measures how far each project is from all the other projects in the network. We formally define the relationship
between the network centrality measures (degree and closeness centrality) and spillovers below.

2.1 An Example Constructing the Patent and Inventor Networks

Before we proceed, Figure 1 below provides a simple example in how to construct the patent and inventor networks in order to make the concepts more concrete. Suppose that there are six inventors and five patents with the following patent-inventor data:

<table>
<thead>
<tr>
<th>Patent 1</th>
<th>Polly &amp; Cindy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent 2</td>
<td>Steve</td>
</tr>
<tr>
<td>Patent 3</td>
<td>Thomas, Elizabeth, &amp; Jack</td>
</tr>
<tr>
<td>Patent 4</td>
<td>Polly &amp; Jack</td>
</tr>
<tr>
<td>Patent 5</td>
<td>Steve &amp; Jack</td>
</tr>
</tbody>
</table>

The first sub-figure in figure 1 shows the two-mode network with both patents and innovators. The second sub-figure shows the “Inventor Network,” where two inventors are connected if they work on a patent together. The third sub-figure is the “Patent Network.” Two patents are connected if they have an inventor in common.

In the inventor network, “Jack” is the most central and he is directly connected to all other inventors except Cindy. In the patent network, both patents 4 and 5 are directly connected to three other patents. Although patents 1 and 3 are not connected, knowledge can indirectly flow between those patents via patent 4. This is because Polly works on both patents 1 and 4, while Jack works on patents 4 and 3.
2.2 A Formal Model for Exploring Network-Mediated Knowledge Spillovers

As discussed, the academic literature has frequently used forward patent citations as a measure of invention quality. Following this convention, we assume that the success level or impact (denoted $S_i$) of each patent “i” is closely related to its count of forward citations, i.e., the citations received from subsequently granted patents. As is typical, we exclude self-citations (both to assignees and to inventors.)

We write:

$S_i = X_i \omega + \varepsilon_i$
where $X_i$ is a vector of observable patent characteristics, $\omega$ is a parameter to be estimated, and $\varepsilon_i$ is an error term.

The FG (2011) model shows how to measure the network ties that could become channels of knowledge spillovers. The model focuses on two network centrality measures: degree and closeness.\(^7\) We define two patents to be linked if they have an inventor in common.

A patent is defined to be from a country if all its inventors are residents of said country, i.e., all inventors have an address in that country on a given patent document. This means, for example, that an Israeli working in the Silicon Valley lab of her multinational employer would be considered “American” for our purposes, because she is a resident of the U.S.

The model assumes that each patent “$i$” may

- receive a positive spillover denoted $\beta$ from all “connected” patents in its country network, and
- that a patent may enjoy positive spillovers from patents that are indirectly connected, but that these spillovers are subject to decay that increases linearly as the distance between the patents in the patent network increases. Formally when the distance between patent $i$ and $j$ is $d(i,j)$, we assume that the success of each patent is $\gamma/\sum_j d(i,j)\) where $\gamma$ is the magnitude of the spillovers.\(^8\)

Under these two assumptions, the success level of each patent $i$ can be written

$$
S_i = X_i \omega + \beta D_i + \frac{\gamma}{\sum_j d(i,j)} + \varepsilon_i,
$$

where $D_i$ is the number of connected patents or degree of the patent.

Formally, closeness centrality is the inverse of the sum of all the (shortest) distances between a focal patent and all other patents multiplied by the number of other patents. Closeness

\(^7\) We define closeness below.

\(^8\) For two patents that are directly connected (that is, share an inventor in common), $d(i,j) = 1$. For two patents that are indirectly linked via a third patent, $d(i,j) = 2$.  

\(10\)
Centrality measures how far each patent is from all the other patents in a network and is calculated as:

\[ C_i = \frac{(N - 1)}{\sum_{j=0}^{N} \frac{d(i,j)}{N-1}}, \]

where \( N \) is the number of patents and \( d(i,j) \) is the distance between Israeli patents \( i \) and \( j \), as measured by the network of coinventions traced out in patent documents. Patents that indirectly link other patents have a higher closeness centrality measure than patents near or at the edge of a network. (See Freeman (1979), pp. 225-226.)

Using (3), the expression for closeness centrality, patent k's success can be rewritten as

\[ S_i = X_i \omega + \beta D_i + \frac{\gamma C_i}{(N-1)} + \epsilon_i. \]

Hence, for each patent (denoted “i”), we calculate the (i) cited patent’s “country network” degree, which is the number of (say) Israeli patents with which the focal patent has a direct link (i.e., an inventor in common) and (ii) the cited patent’s “country network” closeness centrality.

This spillover specification is simple but quite general. The direct spillover from connected patents is \( \beta + \gamma \), while the spillover from patents connected through another (single) patent is \( \gamma/2 \). For patents connected via two other patents, the spillover is \( \gamma/3 \).

When \( \beta=0 \) and \( \gamma=0 \), there are no spillovers. When \( \beta>0 \) and \( \gamma=0 \), there are only direct spillovers. When \( \beta=0 \) and \( \gamma>0 \), there are both direct and indirect spillovers which are exclusively measured by the patents’ closeness centrality. Thus, the theoretical model shows that spillovers depend on the network structure and that they can be measured by constructing the network linking the patents.

By construction, we only consider the possibility of intranational knowledge spillovers, because our networks are based on co-inventions between inventors who “meet” in the same national territory.

Importantly, we address the endogeneity issue associated with network formation. In particular, we define the network to be the “ex-ante” network that was in effect when the
application for the patent was filed. Thus, there is a different network for each patent. We discuss this in detail below.

In order to enable explicit international comparisons of the relative importance of network-mediated knowledge spillovers across different countries, we will construct networks for patents in the relevant classes for countries with large numbers of such patents (Israel, Korea, Japan, Taiwan, Finland, Canada and Germany). We will include this analysis in future versions of the paper.

3. Data and Empirical Work

3.1 Defining and Delimiting Our Patent Populations

We now turn to our empirical work. In order to begin, we need to define the relevant patent classes. From detailed examination of United States Patent and Trademark Office (USPTO) patent class descriptions, we were able to determine the patent classes relevant for information security innovations, broadly defined. These ICT patent classes are shown in Appendix B.

We then collected data from the USPTO on all patents granted in the relevant patent classes. In this data set, we know the number of forward citations, backward citations (citations made to previously granted patents), grant year, application year, location of inventor (hence we know whether the inventor(s) are Israeli), patent class and subclass, patent title and abstract, number of inventors, and the assignee (owner) of the patent.

The number of U.S. patents by country in the relevant patent classes for the years 1985-2014 is given in Table 1. Since there were relatively few information patents in general in these patent classes before 1985, we start with that grant year. In the 1985-2014 period, the USPTO issued approximately 340,000 patents in these patent classes in which all inventors are from the same country. The table shows that more than 50% of the patents were issued between 2005-2014.

---

9 See https://www.uspto.gov/web/patents/classification/uspc726/defs726.htm, accessed 25 June 2017. We included class 709, which does not appear as a relevant patent class in the USPTO document, but, according to research by Arora and Nandakumar (2012), should be included in the information security sector. Nothing changes if we eliminate that class.

10 Patents with missing data account for less than 5% of all patents (and 3% for Israel.)
Because we construct the patent network (for each patent) at the time the patent was applied for, we need to have a large enough existing giant component of connected patents already in existence when a new patent is applied for. We choose 500 as the minimum size of the existing giant component. In the case of Israel, this means we can include patents that were applied for beginning in 2007. In our database, we have patents issued through 2014 and citations through 2016.

In the case of Israel, complete data exist for 876 USPTO patents with Israeli inventors in this period. That is, for these patents, all inventors had an address in Israel. We exclude patents with both Israeli inventors and inventors from other countries (primarily the US) from the main analysis, since we want to focus on the local network.

As noted, we restrict ourselves to patents where all inventors have addresses that place them as resident in the country of interest. No other country has as high a fraction of foreign co-inventors as Israel. Most other countries have very few foreign co-inventors.

The number of Israeli patents is small relative to the total number of such patents. Table 2 shows that Israeli patents as a proportion of all patents granted by the USPTO in these classes increased steadily over the 1985-2014 period, but remained a small percentage of the total. The conventional wisdom regarding Israeli patents in these classes is that they stand out in terms of quality rather than quantity.11

3.2 Construction of the Patent Network

We construct the network of Israeli patents by defining two patents to be linked if they have an inventor in common. Thus, we link patents via the recorded names of inventors. Although the USPTO data are reasonably thorough, the empirical literature has noted the challenges that arise in the "disambiguation" of similar names (Trajtenberg et al., 2009; Ventura, Nugent, and Fuchs, 2015; Marx, Singh, and Fleming, 2015). For the purposes of our study, we think of the

11 It is also possible – and, in fact, likely –that our data include many patents that are not information security patents, strictly defined, and that the Israeli share of a more narrowly defined set of information security patents would be much higher. We chose to err on the side of being reasonably comprehensive in our definition of information security patents.
use of recorded inventor names in USPTO data as raising two main issues, which we refer to as "false positives" and "false negatives."

A **false positive** means that we identify a connection between two patents in the coinvention network, where this connection does not actually exist. A false positive occurs if two (or more) separate inventors have the same name, and we therefore infer more coinventions than actually take place. In order to reduce the potential for false positives, we drop inventors with 100 patents or more patents. Inventor names with a very large number of patents attached to them could, in fact, reflect multiple inventors, and inclusion of such inventors could lead to substantial measurement. In the case of the Israeli network, we individually examined the names of all patent holders with more than 20 patents – and did not find a single case of a false positive. We are thus confident that our results are not driven by false positives in the Israeli data.

A **false negative** means we do not find a connection between two patents due to different spelling, or typing mistakes of the inventors’ names. In order to reduce the probability of false negatives, we standardize all inventor names in the following ways:

1. We use only lower case letters for the names
2. We remove leading and following spaces.
3. We replace all "-" symbols with spaces between names.
4. We remove all punctuation symbols, such as parenthesis, commas etc.

This standardization should help minimize the false negatives in our data. To the extent that they remain, and that our network of coinventions omits important connects, we are underestimating the extent of the network and therefore the knowledge spillovers that may flow through them.

Like many empirical networks, the network of Israeli patents includes one large connected component and many, much smaller components. We refer to the large component as the “giant component.” Closeness is not defined for patents in different components. Since we want to test for both direct and indirect spillovers, in the econometric work, we restrict attention to the

---

12 We note, however, that the qualitative nature of our results is not affected whether we retain or drop inventors with more than 100 patents. There are no such inventors in the Israeli network in any case.
giant component and to patents applied for beginning in 2007. This leaves us with 876 patents in the Israeli giant component.13

We follow a similar procedure in constructing patent networks for the other countries (other than the U.S.) that have generated large amounts of patents in these classes.

The variables used in the analysis are:

- Number of Forward Citations “no self-citations” (excluding forward citations from the same inventor and same assignee)
- Grant Year
- Number of Backward Citations received by the Patent
- Number of Inventors on the Patent
- Degree
- Closeness
- Whether the assignee is in the US

Descriptive Statistics for the Israeli network appear in Table 3.

Israel is unique among countries in that many of its patents have US assignees. Fully 47% of the 876 Israeli patents in the giant component that were applied for beginning in 2007 have US assignees.14 For comparison, no other country has more than 17% “US Assignees” in these patent classes (applied for beginning in 2007,) and most have less than 5% US assignee patents. Hence, in this measure, Israel is “off the charts.”

3.3 Measuring Spillovers via Connected Networks

In this section, we estimate equation the FG (2011) model by estimating equation (4) which we repeat below:

\[ S_i = X_i \omega + \beta D_i + \gamma \frac{c_i}{N-1} + \epsilon_i \]

Recall that \( S_i \), the number of forward citations received by a given patent, is our measure of quality. We exclude self-citations and citations made by patents from the same assignee and the same inventor. We further assume that the number of forward citations received by patent

---

13 Recall that Closeness is not defined for patents in different components.
14 Since the data are from the USPTO, we know whether the assignees are US or foreign entities. In the case of Israel, virtually all non-Israeli assignees are US assignees.
i depends on a vector of observable factors, denoted \( X_i \). These include characteristics of the patent and characteristics of the firm holding the patent (Assignee). \( D_i \) is the *degree* of patent \( i \) in the Israeli network, \( C_i \) is the *closeness centrality* of patent \( i \) in the Israeli network and \( \beta \) and \( \gamma \) are the parameters associated with these centrality measures.

The patent networks play a dual role in expanding the number of citations received by a given patent:

- First, patent networks, as measured by degree and closeness, provide the inventors of a given patent access to useful knowledge that enhances the quality and value of invention \( i \), and hence lead to more citations.
- Second, after invention \( i \) is generated, the network propagates knowledge of this useful invention (and the technical innovations it contains) to other inventor teams working on related technologies, leading to more citations over time.

We are interested in the first effect. Fortunately, we can disentangle these separate effects by constructing a network for each patent at the time the patent was applied for. Although this makes the empirical work computationally intensive, it is necessary in order to examine our main question, which is whether high-quality inventions benefit from the network that was in place when the patent application was filed. To the best of our knowledge, no one has used this methodology when constructing networks.

3.4 Testing for Direct and Indirect Network-Mediated Knowledge Spillovers

Citations are highly skewed; additionally, some of the independent variables (like degree and number of inventors) are also highly skewed. Hence, it makes sense to use logarithms and employ the log/log specification.\(^{15}\) The term “\( \ln \)” before the variable means natural log. The dependent variable used in the regressions in Table 4 is the natural log of forward citations excluding citations from the same inventor and assignee.

The independent variables are the number of inventors on each patent, the number of backward citations, and the degree of the patent, where degree is the number of patents with which the relevant patent has an inventor in common. We control for grant year in every regression.

---

\(^{15}\) We use \( \ln(\text{Forward Citations} + 1) \), since some of the patents do not have any forward citations. Similarly, we use \( \ln(\text{degree}+1) \) and \( \ln(\text{Backward Citations} + 1) \) since degree and backward citations can also take on the value zero. An alternative negative binomial specification gives quite similar results.
3.5 Testing for Direct and Indirect Network-Mediated Knowledge Spillovers: Israel

Column 1 in Table 4 shows the results for the Israeli patents. The estimated coefficient on closeness (\(\gamma\)) is positive and significant (0.20, \(t=3.70**\)), suggesting that there are both direct and indirect knowledge spillovers from “connections” in the giant component. Recall that the coefficient on closeness (\(\gamma\)) captures both direct and indirect spillovers. The estimate for \(\beta\) is -0.04, which suggests that the direct spillover (\(\beta+\gamma = 0.16\)) is less than twice the spillover between patents with one degree of separation (\(\gamma/2=0.10\)).

In columns 2 and 3, we repeat the analysis in column 1 for US and Israeli assignees separately. We find that the estimated coefficient on closeness (\(\gamma\)) is positive and significant for both groups (estimated coefficient =0.20, \(t=2.36**\) for Israeli assignees, coefficient =0.30, \(t=4.58***\) for US assignees,), again suggesting that there are both direct and indirect knowledge spillovers from “connections” in the giant component. The estimates for \(\beta\) are again negative (-0.06 for Israeli assignees, -0.04 for US assignees,) which suggests that the direct spillover is less than twice the spillover between patents with one degree of separation for both Israeli and US assignees.

The estimated coefficient on backward citations is positive and significant in all cases, while the estimated coefficient on the number of innovators positive and statistically significant for the full sample and for “Israeli assignees.”

3.6 Examining the Dual Roles of Patent Networks

Recall that the patent networks play a dual role in expanding the number of citations received by a given patent:

- Patent networks, as measured by degree and closeness, provide the inventors of a given patent access to useful knowledge that enhances the quality and value of invention \(i\), and hence lead to more citations.

- Second, after invention \(i\) is generated, the network propagates knowledge of this useful invention (and the technical innovations it contains) to other inventor teams.
In section 3.5, we measured the first effect, which is our main question. Here, we briefly examine the results when we do not entangle these two effects. To do this, we calculate the network at the end of the data; hence, the network size is the same for each patent in the giant component. We do the analysis for the same 876 patents used the regression column 1 of Table 4. The results appear in column 4 in Table 4.

In this case, we find the estimated coefficient on closeness ($\gamma$) is positive and significant ($0.24, t=2.57^{**}$) and that the estimate for $\beta$ is 0.00 ($t=0.08$). Both $\gamma$ and $\beta$ are larger than in the specification in column 1, suggesting that the second role of patents (propagating knowledge of a useful invention) is important as well.

### 3.7 International Comparisons of Network-Mediated Knowledge Spillovers

In this section, we will estimate network spillover effects in leading countries in their respective giant components. In the analysis, we focus on non-US countries, with at least 500 patents in their giant component beginning in 2007 – South Korea, Japan, Taiwan, Canada, Finland, Germany, Sweden, and France. We will explore the other countries in the next draft of the paper.

### 4. Conclusions and Next Steps

For nearly a quarter century, researchers have used patent citation data to trace out knowledge spillovers across inventions, organizations, and regions. From the inception of this literature, researchers have recognized the potential importance of direct interaction between inventors, but relatively few studies have sought to measure inventor networks explicitly, and fewer still have sought to quantify the degree to which these networks function as mechanisms for the transmission of knowledge spillovers.

Drawing inspiration from related work on open source software projects, this study seeks to advance the literature by using the pattern of inventor interaction traced out in patent documents to create measures of inventor networks; we go on to empirically measure the association between the location of a patent within this network and the quality of invention as measured by forward citations. We apply these techniques in an interesting context – ICT/ information

---

16 We exclude the US from this analysis.
security technology. This is a domain in which Israeli inventors have recently emerged as globally important creators of new technology. Industry accounts suggest that the rapid rise of Israeli firms to this position of global prominence has been driven, in part, by the unusually tight networks that characterize Israeli inventors operating in this domain. These networks allegedly help produce better inventions, and then rapidly convey the new technologies embodied in these inventions to subsequent inventor teams. Despite wide acceptance of this conventional wisdom, no empirical research has yet convincingly related Israeli invention quality to Israeli inventor networks.

This paper presents empirical evidence supporting and extending this conventional wisdom. We find that the quality of Israeli inventions is systematically related to the location of these patents within the Israeli invention network.

These initial results suggest a number of potentially useful directions for further research. While network ties among inventors appear to be strongly correlated with invention quality in Israel, we still know little about the genesis of these ties. Conventional wisdom points to the importance of military service within elite groups like Unit 8200, but no large-sample statistical study has formally tested this popular belief. However, it is possible, in principle, to measure the importance of veterans of Unit 8200, and other elite Israeli Defense Force units, as central nodes within these networks. Increasingly, veterans openly acknowledge their prior ties to these once secret units, and even list their service as a professional credential on social networks like LinkedIn. In future work, we will seek to use these data to probe the importance of the Israeli military as a source of network ties and a driver of invention quality.

Rapid development of machine learning and text mining techniques, applied to patent data, provide another interesting path forward. Gandal, Naftaliev, and Stettner (2017) were able to separately measure the network connections between inventors (and projects) as well as the movement of specific bits of software code across open source projects, and could therefore recognize particular techniques and technologies, as revealed by the text of patent documents, allowing us to track the movement and evolution of these ideas across patents, in both space and time. This would provide a measure of knowledge flows that is independent of the network, but plausibly influenced by it, allowing for a richer and more direct test of the idea that denser networks really do enhance the diffusion and evolution of useful knowledge.
Finally, our measures of network density are deliberately designed to be time invariant in this paper, but the reality is that the inventor and patent networks evolve over time in ways that we can track in our data. Allowing the networks to evolve temporally may enable us to better distinguish between the idea that denser networks create better ideas from the notion that better ideas create a denser network. As is usually the case in economics, much remains to be done.
References


Appendix A: Relevant Patent Classes for Information Security:17

326, Electronic Digital Logic Circuitry, subclass 8 for digital logic circuits acting to disable or prevent access to stored data or designated integrated circuit structure.

340, Communications: Electrical, subclasses 5.2 through 5.74, for authorization control without significant data process features claimed, particularly subclasses 5.22-5.25 for programmable or code learning authorization control; and subclasses 5.8-5.86 for intelligence comparison for authentication.

365, Static Information Storage and Retrieval, subclass 185.04 for floating gate memory device having ability for securing data signal from being erased from memory cells.

380, Cryptography, subclasses 200 through 242 for video with data encryption; subclasses 243-246 for facsimile encryption; subclasses 247-250 for cellular telephone cryptographic authentication; subclass 251 for electronic game using cryptography; subclasses 255-276 for communication using cryptography; subclasses 277-47 for key management; and subclasses 287-53 for electrical signal modification with digital signal handling.

455, Telecommunications, subclass 410 for security or fraud prevention in a radiotelephone system.

704, Data Processing: Speech Signal Processing, Linguistics, Language Translation, and Audio Compression/Decompression, subclass 273 for an application of speech processing in a security system.

705, Data Processing: Financial, Business Practice, Management, or Cost/Price Determination, subclass 18 for security in an electronic cash register or point of sale terminal having password entry mode, and subclass 44 for authorization or authentication in a credit transaction or loan processing system.

708, Electrical Computers: Arithmetic Processing And Calculating, subclass 135 for electrical digital calculating computer with specialized input for security.

709, Electrical Computers and Digital Processing Systems: Multicomputer Data Transferring, subclass 225 for controlling which of plural computers may transfer data via a communications medium.

710, Electrical Computers and Digital Data Processing Systems: Input/Output, subclasses 36 through 51 for regulating access of peripherals to computers or vice-versa; subclasses 107-125 for regulating access of processors or memories to a bus; and subclasses 200-240 for general purpose access regulating and arbitration.

711, Electrical Computers and Digital Processing Systems: Memory, subclass 150 for regulating access to shared memories, subclasses 163-164 for preventing unauthorized memory access requests.

713, Electrical Computers and Digital Processing Systems: Support, subclasses 150 through 181 for multiple computer communication using cryptography; subclasses 182-186 for system access control based on user identification by cryptography; subclass 187 for computer program modification detection by cryptography; subclass 188 for computer virus detection by cryptography; and subclasses 189-194 for data processing protection using cryptography.

714, Error Detection/Correction and Fault Detection/Recovery, subclasses 1 through 57 for recovering from, locating, or detecting a system fault caused by malicious or unauthorized access (e.g., by virus, etc.).

726 Protection of data processing systems, apparatus, and methods as well as protection of information and services.

Tables

Table 1: ICT Information Economy patents by Country for 2013-2014

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Patents</td>
<td>Share of Patents</td>
</tr>
<tr>
<td>South Korea</td>
<td>4113</td>
</tr>
<tr>
<td>South Korea</td>
<td>7096</td>
</tr>
<tr>
<td>Taiwan</td>
<td>5286</td>
</tr>
<tr>
<td>Japan</td>
<td>6801</td>
</tr>
<tr>
<td>Canada</td>
<td>1057</td>
</tr>
<tr>
<td>France</td>
<td>497</td>
</tr>
<tr>
<td>Germany</td>
<td>2077</td>
</tr>
<tr>
<td>USA</td>
<td>10321</td>
</tr>
<tr>
<td>OTHER COUNTRIES</td>
<td>30007</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>39,711</strong></td>
</tr>
</tbody>
</table>

Note: The table presents the number of patents that originated in the respective country, between 2013 and 2014, and are listed in the USPTO database. We identify a patent as one that was originated in a specific country if all its inventors have addresses listed under that country, according the USPTO data.

Table 2: ICT Information Economy patents 1985-2014

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># of All Patents</td>
<td># of US Patents</td>
</tr>
<tr>
<td>1981-1985</td>
<td>1125</td>
<td>32</td>
</tr>
<tr>
<td>1986-1990</td>
<td>1664</td>
<td>71</td>
</tr>
<tr>
<td>1991-1995</td>
<td>5640</td>
<td>376</td>
</tr>
<tr>
<td>1996-2000</td>
<td>5250</td>
<td>335</td>
</tr>
<tr>
<td>2001-2005</td>
<td>5270</td>
<td>980</td>
</tr>
<tr>
<td>2006-2010</td>
<td>5190</td>
<td>250</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>39,516</strong></td>
<td><strong>4,411</strong></td>
</tr>
</tbody>
</table>

Note: Column 1 presents the number of patents issued by all countries and was granted at each five-year period, between 1985 and 2014. Column 2 presents the number of US patents that were granted at the same period. Column (3) shows the percentages of the total patents that were granted in the same period. We identify a patent as one that was originated in a specific country if all inventors have addresses listed under that country, according the USPTO data.

25
Table 3: Descriptive Statistics Israel

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward Citations</td>
<td>878</td>
<td>2.70</td>
<td>6.87</td>
<td>0</td>
<td>68</td>
</tr>
<tr>
<td>Forward Citations No self citations</td>
<td>876</td>
<td>3.43</td>
<td>5.72</td>
<td>0</td>
<td>68</td>
</tr>
<tr>
<td>Grant Year</td>
<td>876</td>
<td>2012.81</td>
<td>1.38</td>
<td>2008</td>
<td>2014</td>
</tr>
<tr>
<td># of inventors</td>
<td>876</td>
<td>2.79</td>
<td>1.49</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Degree</td>
<td>876</td>
<td>75.45</td>
<td>25.77</td>
<td>0</td>
<td>160</td>
</tr>
<tr>
<td>Backward Citations</td>
<td>876</td>
<td>56.35</td>
<td>33.96</td>
<td>0</td>
<td>247</td>
</tr>
<tr>
<td>Closeness(Cl-1)</td>
<td>876</td>
<td>0.000016</td>
<td>0.000011</td>
<td>0.000012</td>
<td>0.000005</td>
</tr>
</tbody>
</table>

Note: The table presents descriptive statistics for all the Israeli patents that are in the Grant component and were issued between 2007 and 2014. Forward citations include the number of citations a patent receives. Forward citations “No self cites” includes all the citations a patent receives, excluding citations made by patents from the same inventor as that were made by that same assignee. Grant year is the year the patent was approved by the USPTO. Number of inventors are the number of inventors listed as the patent inventors. Degree is the number of patents that are directly connected to the patent in the inventors’ network, as the network formed until the patent application year. Closeness is the patent degrees centrality measure, in the patent network formed until the patent application year. Backward citations are the number of patents that were cited by the patent. We identify a patent as one that was originated in Israel if all its inventors’ home addresses were listed under Israel according the USPTO data.
Table 4: Variant Vs. Invariant Israeli Network

<table>
<thead>
<tr>
<th></th>
<th>(1) Network at the time the patent applied for</th>
<th>(2) Network at the time the patent applied for - US Assignees</th>
<th>(3) Network at the time the patent applied for - Israeli Assignees</th>
<th>(4) Network at the end of the data</th>
</tr>
</thead>
<tbody>
<tr>
<td>In(Degree)</td>
<td>-0.944**</td>
<td>-0.907**</td>
<td>-0.683*</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.006)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>In(Closeness)</td>
<td>0.280**</td>
<td>0.286**</td>
<td>0.204**</td>
<td>0.208**</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.057)</td>
<td>(0.095)</td>
<td>(0.097)</td>
</tr>
<tr>
<td>ln(# of Inventors)</td>
<td>0.081**</td>
<td>0.076</td>
<td>0.162**</td>
<td>0.170**</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.064)</td>
<td>(0.064)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>ln(Backward Cites)</td>
<td>0.191***</td>
<td>0.062**</td>
<td>0.150***</td>
<td>0.007***</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.046)</td>
<td>(0.027)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>N</td>
<td>476</td>
<td>414</td>
<td>461</td>
<td>876</td>
</tr>
<tr>
<td>Adj. R²</td>
<td>0.370</td>
<td>0.307</td>
<td>0.365</td>
<td>0.366</td>
</tr>
<tr>
<td>Grant Year Dummies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note: The dependent variable is the natural log of one plus the number of forward citations. While counting forward citations, we exclude citations made by the patent’s inventors other patents, and citations made by other patents that are listed under the patent’s assignee. Number of inventors is the number of inventors listed in the USPTO data. Backward Cites is one plus the number of citations made by the patent. Degree is the number of patents which are connected to the patent in the Israeli patent network that existed up until the patent application year. Closeness is the closeness centrality measure the patent have in the relevant Israeli patent network. In column (1) we regress on all Israeli patents applied for after 2007 and we define the Degree and Closeness of a patent by looking at the network that existed up until the patent application year. In Column (2) we restrict our sample only to patents with US assignees. In Column (3) we restrict our sample to patents with Israeli assignees. In column (4) we again regress on all Israeli patent that were applied after 2007 and we define Degree and Closeness of a patent by looking at the network at 2014. Standard errors are in parenthesis.

* = significant at 10% level, ** = significant at 5% level, *** = significant at 1% level.