Network-Mediated Knowledge Spillovers in ICT/Information Security
(Draft - Not for Circulation/Citation)

June 1, 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 develops a method to examine and measure the incidence and nature of knowledge flows mediated by the collaboration networks of inventors. We apply the methodology to the information and communication technology (ICT) and information security sectors for the case of Israel, which according to the "urban legend" should have such network-mediated knowledge spillovers.

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. We are especially grateful to Tim Bresnahan for very helpful comments and suggestions that significantly improved the paper. We also thank Eugene Kandel and seminar/conference participants at the 10th Paris conference on Digital Economics, Collegio Carlo Alberto, Hebrew University, UC-Berkeley and Stanford for helpful comments and suggestions.

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

[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 in Israel, a leading force in the ICT industry. Like the other papers in this literature, we assume that success level or impact of a patent is closely related to its count of forward citations.

In order to apply the FG 2011 model, we have to address the issue that patent networks form sequentially and therefore play a dual role in expanding the number of citations received by a given patent:

- First, existing patent networks, as measured by closeness at the time of a patent application, provide the inventors of a given patent access to useful knowledge. This enhances the quality and value of invention *i*, and hence leads to more citations. We refer to this effect as the “ex-ante” knowledge spillover.
- 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 refer to this as the “ex-post” knowledge spillover.

In order to measure both of these effects, we modify the FG 2011 to take account of the network formation in a way that enables us to measure both of these spillovers. To the best of our knowledge, we are the first to both identify and measure these effects

Using data from U.S. PTO patent grants in information security, we find that the quality of the Israeli ICT/information security inventions is systematically linked to the structure of the collaborative network in the case of Israel. In particular, we find a positive and significant “ex-ante” knowledge spillover for the case of Israel. We find that there is an “ex-post” spillover over as well. From our estimates, the ex-ante” spillover is twice as large. This suggests that there are knowledge spillovers in the Israeli network, which improve the quality of patents, as measured by the number of citations. Interestingly, in the brief international comparison that we do at the end of the paper, we find that there is no “ex-post” spillover for other countries.

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

[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.

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 inventors 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 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 ascendancy 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

---

[3] See "The Bloomberg Innovation Index", http://www.bloomberg.com/graphics/2015-innovative-countries/ (accessed 17/12/2016) and "Global Competitiveness Report 2015-2016 - Reports - World Economic Forum", http://reports.weforum.org/global-competitiveness-report-2015-2016/economies/#economy=ISR (accessed 17/12/2016.)

[4] "Companies in Israel – Nasdaq.com", http://www.nasdaq.com/screening/companies-by-region.aspx?region=Middle+East&country=Israel (accessed 17/12/2016)

[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/

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 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 with significant numbers of patents in this sector has more than 17 percent US assignees, and most of the countries have less that 5 percent or fewer US assignees.

We now briefly examine ICT/ information security patents by patent class for several countries and for the state of California, which is considered to be on the forefront of knowledge is ICT/Information Security (as well as other areas.) The percent of patents in each of the ICT/Information Security patent classes is shown in Table 1 for California, Israel, Japan, and Korea. (All tables and figures are in the Appendix.)

[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 athttp://www.israel21c.org/high-tech-elites-to-nurture-arab-israeli-startups/

In the case of Korea, almost 60% of the patents are from categories 365 and 455. These are the two largest categories for Japan as well and account for 36% of the patents. In the case of Israel and California, these percentages are 21% and 24% respectively. Excluding class 455, which is very broad and has a large number of patents, the largest three patent classes for both Israel and California are classes 709, 711, and 714. Patent class 709 covers Electrical Computers and Digital Processing Systems: Multicomputer Data Transferring: Patent Class 711 covers Electrical Computers and Digital Processing Systems: Memory: Patent Class 714 covers the Error Detection/Correction and Fault Detection/Recovery. These classes are more oriented to software than patent classes 365 (Static Information Storage and Retrieval) and Patent Class 455 (Telecommunications,) which are more oriented towards hardware.

When we look at the percent of patents in the "700 classes," less the percent of patents in the other classes containing ICT/information security patents, we see an interesting bifurcation. Israel, Canada, and California have many more patents in the "software" classes, while Korea, Taiwan, and Finland have many more patents in the "non" 700 classes containing ICT/ information security patents. Germany, France, and Japan are in the middle. See Table 2.

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 when there are project spillovers that decrease with decay, 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 *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:

	Inventors
Patent 1	Polly & Cindy
Patent 2	Steve
Patent 3	Thomas, Elizabeth, & Jack
Patent 4	Polly & Jack
Patent 5	Steve & Jack

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.

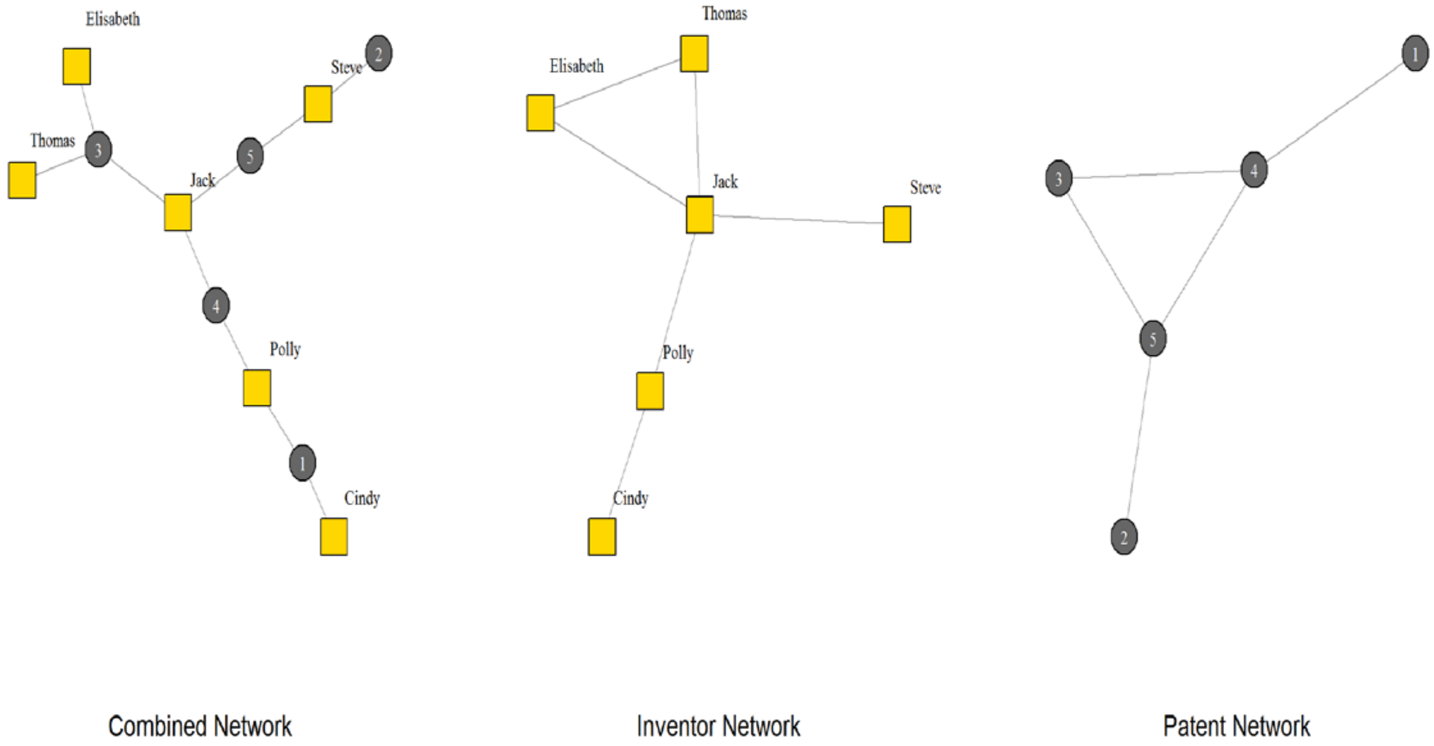

Figure 1: A Two-mode Network and Corresponding Patent Networks

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:

$$(1)\ S_i = X_i\,\omega + \varepsilon_i$$

where X_i is a vector of observable patent characteristics, ω is a parameter to be estimated, and ε_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 the centrality measure closeness, which we define formally below. 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 enjoy positive spillovers from patents that are directly connected and patents that are indirectly connected, but that these spillovers are subject to decay that increases 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 γ is the magnitude of the spillover.[7]

Under this assumption, the success level of each patent *i* can be written

$$(2) \qquad Si = Xi\omega + \frac{\gamma}{\Sigma_j d(i,j)} + \varepsilon i.$$

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 centrality measures how far each patent is from all the other patents in a network and is calculated as:

$$(3) \qquad C_i \equiv \frac{(N-1)}{\sum_{j \in N} d(i, j)},$$

[7] 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.

where N is the number of patents and $d(i,j)$ is the shortest 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

$$(4) \qquad \mathrm{Si} = \mathrm{Xi}\omega + \frac{\gamma \mathrm{Ci}}{(\mathrm{N}-1)} + \varepsilon\mathrm{i}.$$

Hence, for each patent (denoted "i"), we calculate the cited patent's "country network" closeness centrality.

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

Importantly, we need to address the endogeneity issue associated with network formation. In particular, in order to measure the "ex-ante" knowledge spillover, for each patent, we 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 can then take a snapshot of our network at the end of our data. In this case, every patent has the same network. By re-running equation with the same network for each patent, we capture both the "ex-ante" and "ex-post effects. Hence, the ex-post effect is simply the residual spillover, i.e., the total spillover less the ex-ante spillover.

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 *i* 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.[8]

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 3. 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]

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 600 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. Figures 1 and 2 shows the formation and development of the Israeli network and its giant component over time.

In the case of Israel, complete data exist for 881 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.

The number of Israeli patents is small relative to the total number of such patents. Table 3 and 4 shows that Israeli patents as a proportion of all patents granted by the USPTO in these classes

[8] 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.

[9] Patents with missing data account for less than 5% of all patents (and 3% for Israel.

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.[10]

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 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.[11] 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.

[10] 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.

[11] 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.

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 giant component and to patents applied for beginning in 2007. This leaves us with 881 patents in the Israeli giant component.[12]

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
- Closeness
- Whether the assignee is in the US

Descriptive Statistics for the Israeli network appear in Table 5.

Israel is unique among countries in that many of its patents have US assignees. Fully 47% of the 881 Israeli patents in the giant component that were applied for beginning in 2007 have US assignees.[13] 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.”

[12] Recall that Closeness is not defined for patents in different components.
[13] 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.

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:

$$(4)\ S_i = X_i\omega + \gamma\frac{C_i}{N-1} + \epsilon_i$$

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 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). C_i is the *closeness centrality* of patent *i* in the Israeli network and γ is the parameter associated closeness.

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

- First, existing patent networks, as measured by closeness at the time of the patent application, 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.

Fortunately, we can disentangle these separate effects by constructing a network for each patent at the time the patent was applied for. Using the existing networks for each patent, we can estimate (4) to measure the "ex-ante" effect. Although this makes the empirical work computationally intensive, it is necessary in order to examine whether inventions benefit from the network that was in place when the patent application was filed.

We then can re-estimate equation (4) at the end of our data to capture the total spillover, i.e., the ex-ante plus the "ex-post" spillover. In this estimation, all patents have the same network, that is, we measure the network at one point in time. We then calculate the "ex-post" spillover by subtracting the "ex-ante" spillover from the total spillover.

Citations are highly skewed; additionally, some of the independent variables (like the number of inventors) are also highly skewed. Hence, it makes sense to use logarithms and employ the log/log specification.[14] The term "ln" before the variable means natural log. The dependent variable used in the regressions in Table 6 is the natural log of forward citations excluding citations from the same inventor and assignee.

3.4 Measuring the "Ex-Ante" Effect

The independent variables are the number of inventors on each patent, the number of backward citations, and closeness of the patent, where we measure closeness at the time when the patent is applied for. We control for grant year in every regression.[15]

Column 1 in Table 6 shows the results for the Israeli patents. The estimated coefficient on closeness (γ) is positive and significant (0.17, t=3.25***), suggesting that there are knowledge spillovers from ex-ante "connections" in the giant component.

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 (γ) is positive and significant for both groups (0.17, t=2.08** for Israeli assignees, coefficient=0.26, t=4.24*** for US assignees,), again suggesting that there are knowledge spillovers from "connections" in the giant component.

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

Collectively, four large American firms (Apple, Google, IBM, and Intel) hold 28 percent of the "Israeli" patents in the data set. In this sense, Israel is very different from all other countries: they have very small percentages of US assignees. When we exclude patents assigned to these major firms, the estimate of γ remains positive and highly significant (0.18, t=2.83***,) and

[14] We use ln(Forward Citations + 1), since some of the patents do not have any forward citations.
[15] When conducting robustness results, we also include dummy variables for patent class. Our main results are unchanged.

similar to that in column 1 in Table 6. Hence, the results are not affected by excluding the very large firms from the analysis.

3.6 Robustness Analysis: Employing Characteristics from the Innovator Network

In addition to the patent network generated by connections among inventors, there is also a related inventor network. Indeed, as we noted, our data forms a two-mode-network: (I) patents and (II) inventors. The two-mode-network can be partitioned into two types of nodes, e.g. patents and inventors. We can then use the two-mode network to construct two different one-mode networks: (i) the patent network and (ii) inventor network.

Here we add the inventor network to the analysis, where, in the inventor network, two inventors are connected if they work together on a patent. The nodes of the inventor network are innovators and the nodes of the patent network are patents.

We can include the inventor network in the analysis in several ways. One way is to include a dummy variable for inventors who are ranked in the top one percent of all inventors in the country in terms of the number of patents the innovator holds. This dummy variable ("Super Star") takes on the value one if the patent has a top one-percent innovator on the patent and zero otherwise. This controls for inventor quality. When constructing the "Star" variable, we make these calculations at the end of time, reflecting the notion that inventor quality is inherent.

Using the top one percent is ideal because in the giant component, roughly half (about 45 percent) of the patents have such an inventor. In the Israeli patent data, 77% of the inventors have one or two patents, while 10% have more than five patents.[16]
It is interesting to examine whether (controlling for network structure) such "stars" affect the success of the patent. We find that in the case of Israel, beyond the effect it has on the network, the presence of such stars does not affect the success of the patent (-0.004, t=-0.11.) The estimate of γ is essentially unaffected by including this variable. The estimated coefficient on

[16] These numbers are very similar to the open source software data employed by Fershtman and Gandal (2011) in the case of open source software. Two percent of contributors in open source projects worked on five or more projects. In the giant component in the open source data, 50 percent of the projects had a contributor who worked on five or more patents, while outside of the giant component, only eight percent had a contributor who worked on five or more projects. Overall, 90% of the contributors in open source software worked on one or two projects.

γ remains positive and statistically significant (0.17, t=3.16***). Hence, this suggests that, controlling for "stars," we again find that there are both direct and indirect knowledge spillovers.

3.7 Examining the Dual Roles of Patent Networks

In sections 3.5 and 3.6, we measured the first (ex-ante) effect. Here, we calculate the total spillover. 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 881 patents used the regression column 1 of Table 6. The results appear in the last column in Table 6.

In this case, we find the estimated coefficient on closeness (γ) is positive and significant (0.24, t=4.24***.)[17] As we would expect, γ is 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, although smaller. Indeed, the ex-ante spillover is 0.17 (from the first column of Table 6,) and the total spillover (from the last column of table 6) is 0.24. This suggests that in the case of Israel, the first "ex-ante" spillover (0.17) is approximately twice as large as the ex-post spillover (0.07=0.24-0.17.)

4. Brief International Examination of Knowledge Spillovers

In this section, we will briefly examine network spillover effects in countries in their respective giant components.[18] In the analysis, we include countries with at least 500 patents in their giant component beginning in 2007. These countries are South Korea, Taiwan, Japan, Canada, Finland, Germany and France.

[17] As new patents are awarded, the giant component grows. Further, some patents not in the giant component in 2007 become part of the component. In the case of Israel, the giant component at the end of the data has 1046 patents. In the analysis we report in the text, we include the 881 patents that were applied for beginning in 2007. Interestingly, when we repeat the analysis of section 3.7 for the full 1046 patents, the estimate of γ is virtually identical.

[18] We exclude the US. This paper is primarily methodical and employs Israeli data. We do not have the tools to deal with the possibility of inventors with the same names in countries other than Israel. In the case of Korea, more than 50% of the population has one of the following three surnames: Kim, Lee, and Park. Hence, this section should be considered a rough back of the envelope calculation. Nevertheless, we thought that it would worthwhile to include it.

Recall that patent networks play a dual role in expanding the number of citations received by a given patent. The first effect (providing the inventors of a given patent access to useful knowledge that enhances the quality and value of the invention) accounted for approximately 2/3 of the total effect in the case of Israel. The "ex-post" effect (after the invention i is generated, the network propagates knowledge of this useful invention to other inventor teams) accounted for 1/3 of the total effect in Israel.

We now repeat this analysis for the other countries by first estimating the regression in the last column of table Table 6 that captures both effects. We then estimate the first regression in column 1 of table 6. The results are reported in the Table 7.

The results in Table 7 suggest that the first effect might be larger than the second effect. More work would, of course, be needed to make that statement more than suggestive.

5. 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 security technology in Israel. 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.

Finally, 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 the movement of specific bits of software code across open source projects, and could therefore separately measure the network connections between inventors (and projects) as well as the movement of specific ideas and techniques across these projects. In principle, text mining and machine learning techniques could 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.

References

Acemoglu, Daron, 2009. *Introduction to Modern Economic Growth*. Princeton University Press, Princeton, NJ.

Agrawal, A., I. Cockburn and J. McHale, 2006. Gone But Not Forgotten: Labor Flows, Knowledge Spillovers and Enduring Social Capital. *Journal of Economic Geography* 6, 5 (2): 571-591.

Ahuja, G. 2000. Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study. *Administrative Science Quarterly* 45 (3): 425–455.

Alcacer, J. and M. Zhao, 2012. Local R&D Strategies and MultiLocation Firms: The Role of Internal Linkages, *Management Science* 58 (4): 739-753.

Almeida, P., J. Song, and G. Wu, 2001. Mobility of Engineers and Cross-Border Knowledge Building: The Technological Catching-up Case of Korean and Taiwanese Semiconductor Firms, in H. Chesbrough and R. Burgelman (eds.) *Research in Technology and Innovation Management*, 7,:57-84.

Arora, A. and A. Nandkumar, 2012. Insecure Advantage? Markets for Technology and the Value of Resources for Entrepreneurial Ventures. *Strategic Management Journal* 33 (3): 231–251.

Ballester, C., A. Calvó-Armengol, Y. Zenou. 2006. Who's Who in Networks. Wanted: The Key Player. *Econometrica* 74 (5): 1403–1417.

Bernstein, J. and M. I. Nadiri, 1988. Interindustry R&D Spillovers, Rates of Return, and Production in High-Tech Industries. *American Economic Review Papers and Proceedings*, 78 (2): 429-434.

Berry, H., 2014. Global Integration and Innovation: Multi-Country Knowledge Generation within MNCs. *Strategic Management Journal* 35 (6): 869-890.

Branstetter, L., 2006. Is Foreign Direct Investment a Channel of Knowledge Spillovers: Evidence from Japan's FDI in the United States. *Journal of International Economics* 68 (2): 325-344.

Branstetter, L., and K. Saggi, 2011. Intellectual Property Rights, Foreign Direct Investment, and Industrial Development. *Economic Journal* 121 (555): 1161-1191.

Branstetter, L., and M. Sakakibara, 2002. When Do Research Consortia Work Well and Why? Evidence from Japanese Panel Data. *American Economic Review* 92 (1): 143-159.

Branstetter, L., 2001. Are Knowledge Spillovers International or Intranational in Scope? Microeconometric Evidence from Japan and the United States. *Journal of International Economics* 53 (1): 53-79.

Calvó-Armengol, A. and M. O. Jackson. 2004. The effects of social networks on employment and inequality. American Economic Review 94 (3): 426–454.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou. 2009. Peer effects and social networks in education. Review of Economic Studies 76 (4): 1239–1267.

Fershtman, C., and N. Gandal. 2011. Direct and indirect knowledge spillovers: the “social network” of open-source projects. The RAND Journal of Economics 42 (1): 70–91.

Freeman, L. 1979. Centrality in social networks: Conceptual clarification. Social. Networks 1 (3): 215–239.

Gandal, N., and U. Stettner. 2016. Network Dynamics and Knowledge Transfer in Virtual Organizations. International Journal of Industrial Organization 48 (C): 270-290.

Gandal, N., P. Naftaliev, and U. Stettner. 2017. Following the Code: Spillovers and Knowledge Transfer. CEPR Discussion Paper, DP11851.

Gomes-Casseres, B., J. Hagedoorn, and A. Jaffe. 2006. Do Alliances Promote Knowledge Flows? *Journal of Financial Economics*, 80(1), 5-33.

Grossman, G. and E. Helpman, 1991. *Innovation and Growth in the Global Economy*. MIT Press, Cambridge.

Irwin, D. and P. Klenow, 1994. Learning-by-Doing Spillovers in the Semiconductor Industry. *Journal of Political Economy* 102 (6): 1200-1227.

Goyal, S., M. J. Van Der Leij, andJ. L. Moraga-González. 2006. Economics: An Emerging Small World. *Journal Political Economy* 114 (2): 403–412.

Griliches, Z. 1979. Issues in Assessing the Contribution of R&D; to Productivity Growth. *The Bell Journal of Economics* l0 (l): 92-116. Also reprinted in E. Wolff, ed., *The Economics of Productivity*, vol. I, Cheltenham: Elgar, 1997, 256-80.

Griliches, Z. 1992. The Search for R&D Spillovers. *The Scandinavian Journal of Economics* 94 (0): S29-S47.

Jackson, M. and L. Yariv. 2007. Diffusion of Behavior and Equilibrium Properties in Network Games. *American Economic Review* 97 (2): 92–98.

Jaffe. A. 1986. Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value. *American Economic Review* 76 (5): 984-1001.

Jaffe, A., Trajtenberg, M., and R. Henderson. 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics* 108 (3): 577-598.

Jaffe, A. and M. Trajtenberg. 2002. *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. MIT Press, Cambridge.

Jones, B. F, 2014. The Knowledge Trap: Human Capital and Development Reconsidered (WP No. 14138). National Bureau of Economic Research.

Karlan, D., M. Mobius, T. Rosenblat, and A. Szeidl. 2009. Trust and Social Collateral. *Quarterly Journal Economy* 124 (3): 1307–1361.

Keller, W. and S. Yeaple, 2013. The Gravity of Knowledge. *American Economic Review* 103 (4): 1414-1444.

Keller, W., 2004. International Technology Diffusion. *Journal of Economic Literature* 42 (3): 742-782.

Keller, W., 1998. Are International R&D Spillovers Trade-Related? Analyzing Spillovers Among Randomly Matched Trade Partners. *European Economic Review* 42 (8): 1469-1481.

M. Marx and J. Singh, L. Fleming. 2015. Regional Disadvantage? Employee Non-compete Agreements and Brain Drain. Research Policy 44, 941-955.

Romer, P., 1986. Increasing Returns and Long-Run Growth. *Journal of Political Economy* 94 (5): 1002-1037.

Romer, P., 1990. Endogenous Technological Change. *Journal of Political Economy* 98 (5) : S71-S102.

Scherer, F. M. 1982. Inter-industry Technology Flows and Productivity Growth. *The Review of Economics and Statistics* 64 (4): 627-634.

Senor, D. and S. Singer. 2009. *Start-Up Nation: The Story of Israel's Economic Miracle*. McClelland and Stewart.

Singh J. 2008. Distributed R&D, Cross-Regional Knowledge Integration and Quality of Innovative Output. *Research Policy* 37 (1): 77-96.

Trajtenberg, M. 1990. *Economic Analysis of Product Innovation – The Case of CT Scanners*. Harvard University Press, Cambridge, MA.

Trajtenberg, M., G. Shif, and R. Melamed. 2009. The Names Game: Harnessing Inventors, Patent Data for Economic Research. *Annals of Economics and Statistics* 93/94: 79-108.

Ventura, S., Nugent, R., and Fuchs, E. 2015. Seeing the Non-Stars: (Some) Sources of Bias in Past Disambiguation Approaches and a New Public Tools Leveraging Labeled Records. *Research Policy* 44 (9): 1672-1701.

Appendix A: Relevant Patent Classes for Information Security:[19]

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.

[19] See https://www.uspto.gov/web/patents/classification/uspc726/defs726.htm, accessed 25 June 2017.