Paper to be presented at DRUID17
NYU Stern School of Business, New York, June 12-14, 2017

Modeling Knowledge Dynamics of the Global Innovation Network Reveals Pre-eminent Technology Domains

MICHAEL Verba
Tilburg University
Tilburg School of Economics and Management
mikey@openmailbox.org

Abstract
In this paper, drawing on techniques from patent metrics, network analysis, and probability theory, we model the global system of innovation as a dynamic network. The sphere of technologically relevant knowledge is conceptualized as a reflexive, directed, link- and node-weighted complex network, with distinct spheres of knowledge (or technology domains) representing network nodes and learning (or knowledge flows) across domains acting as inter-nodal links. We instantiate the nodes of the knowledge network from patent categories. Links between technology domains, representing knowledge transfer between fields of technology, are constructed from patent citations. The empirical knowledge network is constructed from a sweeping patent database, including records from more than 100 patent-granting authorities over the 22-year period spanning 1991-2012. After establishing the structure of the global innovation network, we simulate its dynamics and study its evolution over time. The modelling exercise reveals technological trends and provides a ranking of technologies in terms of their level of technological dynamism.

Jelcodes:O31.D83
Learning Hubs:  
Modelling Knowledge Dynamics of the Global Innovation Network Reveals  
Pre-eminent Technology Domains  


JEL: O31, O33, D83, D85  
Keywords: Global system of innovation; knowledge space; knowledge network; network dynamics; technological change; innovation system; patents; citations.

“Knowledge is in every country the surest basis of public happiness.”  
George Washington (January 8, 1790)  

“The mechanisms and dynamics driving the growth of public technological knowledge, however, are still poorly understood.”  
Carnabuci and Bruggeman (2009, p. 608)

I. Introduction

In a work that gave birth to the term “knowledge economy,” Fritz Machlup (1962) presented evidence that in modern societies knowledge-intensive industries constitute an increasing share of total economic activity. Whether measured by employment in knowledge-intensive sectors, by R&D performed, or by patents granted, indicators show that knowledge is a quantitatively increasing—and increasingly critical—part of the globalized economic system. Not only is the share of knowledge-related industries in the economy increasing, so is their importance for future growth.

Growth theories suggest that long-term growth is only possible with continued creation of technologically relevant knowledge. In endogenous theories of growth, technological knowledge serves a dual role, as a factor in production, as well as an autocatalyst—an input necessary for its own creation. The possibility of infinite growth then, partly rests on the strength and direction of the self-generating aspect of knowledge. The more important the share of knowledge in economic production, and the more indispensable its role, the greater the need to understand its structure, features, and dynamics.

Yet there is much about knowledge processes that we still do not know (Carnabuci and Bruggeman, 2009). Particularly in macro-level modes of analysis, more can be done to distinguish between types of knowledge bases and reveal their patterns of interaction. Furthermore, in a World moved by technological trends, where novelties fade away as quickly as they emerge, the question of the relative importance of different technologies arises in policy and management settings.

The Twentieth Century has been called, alternately: “the Electronic Century,” “the Pharmaceutical Century,” “the Space Age,” and “the Nuclear Age,” and similar stamps exist for every decade. But when it comes down to it, among the extant or ascendant technologies of the time, which
is the most important and can be called the defining technology of the decade? Are we living in a decade where innovation is dominated by Biotech, wireless communication, motor vehicles, or Nanotech? Which is more essential to the functioning of the global innovation system? At any given moment in time how do different technologies compare in terms of their importance and can a cardinal or ordinal ranking of importance be constructed? In this study we outline a methodology for approaching these kinds of questions and provide a first sketch of answers.

Decomposition of knowledge stock into more than 600 categories makes it possible to glean a high-resolution view of the state and temporal evolution of the global system for the production of technological knowledge. And by modeling knowledge dynamics in a way that takes into account interaction between distinct knowledge bases, we are able to measure the contribution of each technological field to the system as a whole, providing an empirical and quantitative measure of each technology’s importance to the pace of global technological progress.

A network consists of a number of elements (also called nodes or vertices) joined by a number of links (also called edges). We conceptualize and model the stock of technological knowledge as a network of distinct but interacting technology domains. To operationalize the knowledge network we rely on findings and techniques from patentometrics, and methods from probability theory and network science. Our analysis exploits a unique global database covering patents granted at more than 100 intellectual property authorities during the 1991-2012 time period and including inter-patent citations. This data is a rich source of information on the state of technology available to humanity. The network is instantiated with patent categories representing nodes, and inter-category patent citations serving as network links.

This article explores the global innovation network, focusing on the way knowledge circulates throughout the network. Analysis of network dynamics provides several advantages in the context of citation networks. Citation networks can be meaningfully interpreted as networks of flows. Numbers of citations between patent categories in a technology network, or between scientific fields, in a bibliometric network, are a measure of the knowledge transferred during a given time period. For knowledge networks, a flow-based analysis is thus truer to the object being studied than a purely topological approach more appropriate for physical infrastructure (such as road) networks, or undirected graphs representing connectivity arrangements (as in telephone circuit or computer networks), where flow volumes may not always be important (as, for example, in network vulnerability analysis).

To model knowledge flows on the global innovation network we apply techniques from network analysis and that facilitate description of network dynamics. We start with the observation that flow on a network can be modeled with a random walk. Just as a charge induces a current on a printed electronic circuit, we use a random walk to generate flux on the topology of the innovation network. The modeling methodology allows us to calculate knowledge flow, that is, the intensity of technological learning within discrete technological fields. It thus provides a way to identify innovation “hot spots” or technology learning hubs characterized by intense acquisition of new technological knowledge.

II. From Knowledge Stock To Patent Network

Several literature streams attempt to place the role of technological knowledge in modern societies in a conceptual-theoretical framework. First, growth theory provides mathematical models that consider the

\[^{1}\text{In this study the term “category” is used in a general sense to indicate a categorization scheme at any level of aggregation: patent section, class, subclass, or some other classification system.}\]

\[^{2}\text{In a companion study we consider in detail the community structure of the global knowledge network as revealed by knowledge flow patterns.}\]
role of knowledge in the economy (Solow, 1956, 1957; Romer, 1990). Second, work on systems of innovation (SI) (Freeman, 1987; Lundvall, 1992; Nelson, 1993) and technological paradigms (Dosi, 1982) has paid considerable attention to technological change. SI has considered the economic role of knowledge in a broad social and institutional context, while the latter has been concerned with the evolution of technology. A third source of what we know about the economic role of knowledge has come from surveys (Mairesse and Mohnen, 2010) and empirical literature on innovation indicators (Archibugi and Coco, 2005; Gault, 2013), including the large body of research based on indicators derived from patent filings, with the work of Griliches (1990) and Jaffe and co-authors among the landmark contributions (Jaffe, 1986; Jaffe, Trajtenberg and Henderson, 1993). Finally, a growing subfield of research on the role of knowledge in the economy is represented by the new literature on knowledge networks (Valverde et al., 2007; Carnabuci and Bruggeman, 2009; Lafond, 2015; Feldman, Kogler and Rigby, 2015; Triulzi, 2015). In the next Subsection we explore a number of advantages from conceptualization of knowledge stock as a knowledge network.

A. Knowledge as a Network

In macro-level literature knowledge tends to be approached as an undifferentiated “stock” of ideas. For macroeconomic models this treatment is not inappropriate, but the inherent limitation of aggregation is that it sweeps into one basket items quite dissimilar. Antonelli, Krafitt and Quatraro (2010) argue for a disaggregative approach to knowledge stock as a means to study “the internal structure” of knowledge varieties.³

One of the largest organism found in nature is the Aspen grove. Although on the surface the grove appears as a forest composed of many trees, below ground it is one entity, with a common root structure. While a story of technological evolution at the level of domains can be told with simple count data and trend analysis, by itself such an exercise would overlook the hypogeal connections between technologies—the exchange of knowledge between technological fields. It would be like considering as insular, individual trees in an Aspen grove, ignoring the network of roots that unites them.

Scholarship on the sociology of technology has used the word “network” as a metaphor in descriptions of technological progress (Podolny and Stuart, 1995). Perhaps part of the appeal of the network metaphor is that on an intuitive level it speaks to the heterogeneous and relational nature of technical knowledge.

At the abstract level, aggregated stock of technologically useful knowledge, or what is sometimes referred to as “knowledge space,” can be conceived as partitioned into distinct technology domains. It may be argued that every idea is unique, but suffice it to acknowledge that technological knowledge is composed of separate technology domains, each with distinct attributes and an unique history and future trajectory. New knowledge draws on pre-existing knowledge in own and other fields of technology. The size of the fields of knowledge, their growth rates, and the pattern of knowledge flows between them characterize the knowledge network.

When it comes to knowledge, the term “network” is more than just a metaphor. Knowledge processes not only lend themselves naturally to network interpretation but can also be studied with the methods of network science. Applied to a knowledge system, a network-analytic approach provides a way to study the state of the separate domains of knowledge, as well as a methodology for revealing the web of interrelationships between domains. Network analysis lets us see the forest, as well as the trees.

³In which the authors argue that “in order to assess the relationship between the generation of new knowledge and economic growth, the focus on knowledge capital stock and traditional indicators of its quality such as patent citations and litigations, is not sufficient to capture the qualitative changes that affect the internal structure of knowledge bases at firm level and at more aggregate levels of analysis” (Antonelli, Krafitt and Quatraro, 2010, p. 51).
B. Once Again, Patents as Indicators

There are two primary methods by which knowledge stock is operationalized empirically as a knowledge network. Both rely on some form of citations to measure knowledge transfer. A bibliometric approach forms nodes from articles, journals, or scientific fields and links from article citations. The second, patent-analytic, approach instantiates the network from categories of patents, with links formed from patent citations. The former represents the body of scientific knowledge and the interactions between scientific fields; the latter provides a way to represent technological knowledge. In this study we rely on the patentometric methodology.

Empirical patent networks rest on foundations from the literature using patents as indicators of technological knowledge, invention and innovation. The same advantages and limitations that are present in the literature using patents as knowledge indicators can be resurrected in the context of patent networks, and we revisit them in brief.

Despite their frequently-acknowledged limitations\(^4\) patent filings are a rich and valuable source of information on the state and development of technology.\(^5\) Buchanan, Packard and Bedau (2011) write that:

“Since the patent record is a fairly complete and accurate history of most technological innovations, mining the patent record is a feasible and empirical way to observe how technology actually evolves.”

Buchanan, Packard and Bedau
(2011, p. 109)

Patent data has been brought to bear on a number of important questions in economics and history of technology. With patent data researchers have tracked the distribution of innovative activity at the level of countries (Furman, Porter and Stern, 2002) and regions (Orlando and Verba, 2005); identified general-purpose technologies (Moser and Nicholas, 2004); traced the progress of specific industries such as information and communication technologies (ICT) (Antonelli, Krafft and Quatraro, 2010), semiconductors (Triulzi, 2015) and printed electronics (Kim, Cho and Kim, 2014); and studied the evolution of technology in general (Buchanan, Packard and Bedau, 2011).

III. Formal Model

In the present section we specify a formal graph-theoretic model of our knowledge network. We model the knowledge network as a node- and edge-weighted, directed, reflexive graph, denoted by \(G = (V, E, h, w)\), where \(V = \{1, ..., n\}\) is a finite set of vertices representing technology domains and \(E = \{1, ..., m\}\), is a finite set of ordered pairs of vertices in \(V\), where each edge \((i, j) \equiv e_{ij}\), represents directed links between technology domains, with \(i\) as the head (source) and \(j\) as the tail (target) of the link. Each vertex \(v_i \in V\) is associated with a node weight \(h_i = h(v_i)\) given by the node-weight function \(h : V \rightarrow \mathbb{R}^+\). In like fashion, each edge \(e_{ij} \in E\) connecting nodes \(i\) and \(j\) is associated with a weight value \(w_{ij} = w(e_{ij})\) representing the connection strength between nodes, given by the weight-edge function \(w : E \rightarrow \mathbb{R}^+\). The adjacency matrix \(A(G) = [a_{ij}]\) of graph \(G\) is a square \(n \times n\) matrix consisting of elements:

\[
 a_{ij} = \begin{cases} 
 w(e_{ij}) & \text{if } e_{ij} \in E \\
 0 & \text{otherwise}
 \end{cases}
\]

In the adjacency matrix \(A\) element \(a_{ij}\) contains the weight of the link from the source

---

\(^4\)The list of limitations includes the fact that not all ideas are patentable, not all ideas are patented, there exist differences between industries in their propensity to patent (Cohen, Nelson and Walsh, 2000), and observed changes in the role and institutional practices within patent systems. The seminal source qualifying the use of patents is Griliches (1990); extended discussion can be found in Archibugi (1992), and a compendary updated overview in Antonelli, Krafft and Quatraro (2010, p. 54).

\(^5\)Acs, Anselin and Varga (2002) and Hagedoorn and Cloodt (2003) are among many studies validating the use of patent counts as an indicator of innovative activity.
node $i$ to the target node $j$. A weight value of 0 is assigned if a link is absent. Links are drawn in the direction of knowledge flow, which is opposite from the direction of knowledge search indicated by the citation.

To fully represent the knowledge network, graph $G$ has two distinguishing attributes: it is reflexive and node-weighted. Unless stated otherwise, the default assumption in network analysis is to exclude the possibility of loops (or self-links), setting $w_{ii} = 0 \forall e_{ii} \in E$. By contrast, our graph is reflexive. Self-loops are allowed because technological domains can draw on internal knowledge. In fact, we show that internal knowledge stock is an important part of the generative process. Hence, values in the diagonal of the adjacency matrix are not restricted to 0, although we do not exclude the theoretical possibility that there may be some technology domains that source knowledge exclusively from other domains. The operative restriction on the weight of self-links is $w_{ii} \geq 0 \forall e_{ii} \in E$.

The second attribute that differentiates our knowledge network from most graphs is that it is node-weighted. Node weights given by the function $h$ represent the size of the domain of knowledge. The vector $H(G) = [h_i]$, defined as:

$$h_i = \begin{cases} h(v_i) & \text{if } v_i \in V \\ 0 & \text{otherwise} \end{cases}$$

contains the sizes of technology domains. The state of the network is completely specified by an $n \times n$ adjacency matrix $A$ and an $n$-dimensional vector of node sizes $H$. Below, we consider our options for implementation of a domain-to-domain knowledge network and outline the empirical strategy we adopt.

IV. Building the Network

A. Data

In this paper we work with patent data covering the 22-year period spanning 1991-2012 drawn from the April 2013 edition of the PATSTAT database.$^6$ The value of working with recent data is that it can answer questions about the latest trends and the current state of technology. We exploit a comprehensive proprietary database containing more than 83 million patent records from 105 patent authorities. Included also are more than 135 million citation records between patents and from patents to non-patent literature.

The database contains patent applications (where available) and grants, including from the three globally important patent offices of EPO, USPTO and the Japan Patent Office. On the whole, the database provides a uniquely comprehensive view of the evolution of technological progress and inter-technological learning up to the most recent years, but we observe two distortions arising from limitations in coverage. First, for grants, the number of records drops towards the end of the coverage period due to the well-known patent processing lag. Second, completeness of citation data is not uniform across all patent offices, but improves over time and peaks towards the end of the Oughties. For more information on data coverage see Subsection F; Section V addresses how the modeling methodology we adopt overcomes the above limitations.

B. Nodes and Links

The International Patent Classification (IPC) System identifies and defines distinct technology categories, which become the nodes of the empirical knowledge network. We use inventor-origin patent citations to construct the links of the network. In this directed and link-weighted network the number of citations from domain $A$ to domain $B$ represents the intensity with which technological inventions in $A$ source knowledge from $B$. The greater the number of citations, the greater is the weight of the link. The network is reflexive, because technological domains can source knowledge from their internal knowledge stock. In fact, at

$^6$Data for years outside this period is extremely sparse and we ultimately exclude it from analysis.
high levels of aggregation internal knowledge is the largest single source of knowledge for most technologies.

The IPC scheme is updated at regular intervals and 20 vintages of the IPC exist at the time of this writing. Our database is linked to the January 2006 version of the IPC (also called Version 8), which is divided into 8 sections, 129 classes, 639 subclasses, 7,314 main groups and 61,397 subgroups. The section level is too aggregated for a fine-resolution view of knowledge dynamics, while the subgroup scheme is not uniformly applied, since patent authorities can opt to classify only to the level of the main group. For our purposes, the subclass is a goldilocks level of aggregation: disaggregated enough to provide a high-resolution view of knowledge dynamics, but not too disaggregated to pose questions about the underlying indicators that had been validated only at an aggregated level.

C. Unique Inventions

To analyze knowledge dynamics between technologies, we take unique inventions granted within each subclass, as well as citations between unique inventions, aggregated at subclass level. Because the patent record can contain multiple documents per invention, reflecting different types of legal events, proper handling of the record base is important. Within the same patent office, the multiple events include (but are not limited to) patent application, granting, modification, inclusion of additional claims through a patent of addition, or a divisional patent that splits the claims to an existing invention into separate bundles of rights. Most importantly, there exist applications and grants covering the same invention in multiple jurisdictions. Inclusion of all records in the construction of the knowledge network would lead to distortive duplication. We overcome this potential hazard by extracting the unique invention from the multitude of redundant records.

A set of related patents pertaining to the same invention is called a “patent family.” Every patent family will contain the earliest patent grant on record, or First Grant, and possibly one or more subsequent grants. To obtain a count of unique inventions within technology domains we can sum the number of patent families, or equivalently, the number of First Grants, linked to each subclass. Table 1 shows the total number of unique inventions in our dataset, cross-tabulated by IPC section.

D. Inventor Knowledge

To identify as closely as possible the state of inventor knowledge at the time of invention, we only include inventor-origin citations placed in the First Grant document. In some cases it may be that the list of citations placed by the inventors diverges between the first grant and subsequent patent filings. Figure 1 illustrates a potential scenario whereby a first grant and a subsequent grant within the same Patent Family diverge in the patents they cite. In Figure 1 circles indicate patents and arrows indicate citations between patents. The circle marked F is the first grant, and the circle labeled S is a patent granted some time later in another jurisdiction. Patents cited in the first-grant only are represented by black circles. The first grant F and the subsequent grant S overlap in one citation, but each cites a patent not cited by the other.

What do these discrepancies represent and how should they affect empirical representation of inventor knowledge? There is no reason to doubt the overlapping citation; it most likely represents inventor learning from the prior invention. There is also no reason to exclude the citation included in the first grant but excluded in the subsequent document. The subsequent exclusion could be a result of any number of alternative scenarios: clerical error, inventor forgetting, or deliberate suppression of the earlier reference. In all of these cases the original citation subsequently excluded would still represent inventor knowledge at the time of invention and we would still want to include it in the construction of the knowledge network. In other words, all citations indicated in the original grant should be included.

The citation excluded in the first grant
Table 1—IPC Sections and Patent Counts

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Section Title</th>
<th>Patents</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Human necessities</td>
<td>1,578,620</td>
</tr>
<tr>
<td>B</td>
<td>Performing operations; transporting</td>
<td>1,881,160</td>
</tr>
<tr>
<td>C</td>
<td>Chemistry; metallurgy</td>
<td>1,355,603</td>
</tr>
<tr>
<td>D</td>
<td>Textiles; paper</td>
<td>161,125</td>
</tr>
<tr>
<td>E</td>
<td>Fixed constructions</td>
<td>431,607</td>
</tr>
<tr>
<td>F</td>
<td>Mechanical engineering; lighting; heating; weapons; blasting</td>
<td>934,943</td>
</tr>
<tr>
<td>G</td>
<td>Physics</td>
<td>2,220,617</td>
</tr>
<tr>
<td>H</td>
<td>Electricity</td>
<td>2,327,220</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>10,890,895</td>
</tr>
</tbody>
</table>

Note: Number of uniquely identifiable inventions receiving patent protection at a national or supra-national patent office during the 1991-2012 period. Utility models and industrial designs are not included. Applications filed under the Patent Cooperation Treaty are also excluded.

but included in a subsequent application, on the other hand, is more difficult to interpret. It is possible that the citation to the patent represented by the gray circle in the Figure, documents knowledge flow that was suppressed in the original application for strategic reasons, but that has been revealed subsequently. It is just as possible, however, that the new citation represents inventor learning after the original application, and consequently, after the invention, perhaps as a result of interaction with the first patent office. In this case the new citation does not represent generative knowledge transfer leading to the invention, and should not be included. Citations of this type are a “mixed bag” (Leydesdorff, 2008) that would only add noise to the indicator of knowledge flow.

Discrepancies between filings within the same patent family can also be influenced by the tendency of patents to cite documents within the same jurisdiction. Patent documents filed in the U.S. tend to cite U.S. patents and patents pertaining to the same invention filed at the EPO tend to cite European patents. If that is the reason behind the appearance of citation C and if C is only a replacement of citation A, representing the same invention, then A and C are interchangeable and it would be fine to include either. However, inclusion of both A and C citations in the total citation count would be in error, since it would count twice the flow of knowledge from the same invention, distorting the weight of the links.

For above reasons, when constructing the knowledge network, we rely only on inventor-origin citations from the First Grant patent. Relying only on the First Grant patent has the added advantage of de-duplicating the patent record.

E. Overview of the Empirical Network

We construct the global knowledge network by calculating the fractional count of the total number of links between IPC subclasses. The timeline covered by the analysis starts in 1991 and ends in 2012. Between 1980 and 1990, data are very sparse, with less than 1000 links annually, so we excluded these years from the analysis. We then split the timeline for which data are available into two periods of 11 years in duration. Network I captures knowledge dynamics during the 1991-2001 period, while Network II covers the years 2002-2012. Table 2 presents the total number of directed links for each year. For Network I, the data are sparse for the first half of the long decade of 1990s, but rises to a level above 100,000 annually that is observed in later years. Although we would prefer to have better documentation of knowledge flows between technology domains for the first years of Network I, knowledge flow estimates should not be affected by the absolute weight of the links, since only relative weights determine the steady-state distribution of flow on the network. Limited network coverage in the 1991-1996 period should not affect our node flow estimates as long as all link weights in Network I,
irrespective of the domains they connect, are under-estimated by the same proportion compared to Network II.

The constructed empirical network pertaining to the 1991-2001 period (Network I) consists of 629 technological sub-classes and 180,218 directed weighted links (including self-links). The network representing knowledge space in the 2002-2012 period (Network II) is only slightly bigger, at 637 nodes, but more dense, with 245,795 links. Networks I and II represent separate cross-sections, or “snapshots,” of the network at two time periods. Analyzing knowledge dynamics on one cross-section can reveal network structure and the relative importance of the various domains of knowledge, at that time period. By comparing knowledge dynamics along the network during different time periods we can learn something about the evolution of the knowledge network.

V. Modeling Network Knowledge Dynamics

A. Knowledge Networks: From Statics to Dynamics

Prior research viewing knowledge systems as complex networks has tended to treat these networks as static graphs, and to use the tools of network science designed for analysis of network topology.\(^7\) But, additional dimensions about the functioning of innovation systems can be gleaned from a dynamics-oriented perspective and modeling methodology. the system of production and dissemination For a number of reasons we find the dynamics-oriented approach more suitable for modeling knowledge networks than exclusively topological

---

\(^7\)Subsection title after Balland, Boschma and Frenken (2015).

\(^8\)By “topology” we mean the “specific pattern of connections” between nodes of the network (Gross and Blasius, 2008, p. 259).
Table 2—Construction of Empirical Network

<table>
<thead>
<tr>
<th>Number</th>
<th>Year</th>
<th>Directed Links</th>
<th>Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1991</td>
<td>1022</td>
<td>I</td>
</tr>
<tr>
<td>2</td>
<td>1992</td>
<td>2584</td>
<td>I</td>
</tr>
<tr>
<td>3</td>
<td>1993</td>
<td>5685</td>
<td>I</td>
</tr>
<tr>
<td>4</td>
<td>1994</td>
<td>10,227</td>
<td>I</td>
</tr>
<tr>
<td>5</td>
<td>1995</td>
<td>19,584</td>
<td>I</td>
</tr>
<tr>
<td>6</td>
<td>1996</td>
<td>22,869</td>
<td>I</td>
</tr>
<tr>
<td>7</td>
<td>1997</td>
<td>43,189</td>
<td>I</td>
</tr>
<tr>
<td>8</td>
<td>1998</td>
<td>75,830</td>
<td>I</td>
</tr>
<tr>
<td>9</td>
<td>1999</td>
<td>110,294</td>
<td>I</td>
</tr>
<tr>
<td>10</td>
<td>2000</td>
<td>125,742</td>
<td>I</td>
</tr>
<tr>
<td>11</td>
<td>2001</td>
<td>131,285</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>548,311</td>
<td>I</td>
</tr>
<tr>
<td>12</td>
<td>2002</td>
<td>137,407</td>
<td>II</td>
</tr>
<tr>
<td>13</td>
<td>2003</td>
<td>146,527</td>
<td>II</td>
</tr>
<tr>
<td>14</td>
<td>2004</td>
<td>153,779</td>
<td>II</td>
</tr>
<tr>
<td>15</td>
<td>2005</td>
<td>153,441</td>
<td>II</td>
</tr>
<tr>
<td>16</td>
<td>2006</td>
<td>142,788</td>
<td>II</td>
</tr>
<tr>
<td>17</td>
<td>2007</td>
<td>141,452</td>
<td>II</td>
</tr>
<tr>
<td>18</td>
<td>2008</td>
<td>139,945</td>
<td>II</td>
</tr>
<tr>
<td>19</td>
<td>2009</td>
<td>118,610</td>
<td>II</td>
</tr>
<tr>
<td>20</td>
<td>2010</td>
<td>101,768</td>
<td>II</td>
</tr>
<tr>
<td>21</td>
<td>2011</td>
<td>71,143</td>
<td>II</td>
</tr>
<tr>
<td>22</td>
<td>2012</td>
<td>22,810</td>
<td>II</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1,320,070</td>
<td>II</td>
</tr>
</tbody>
</table>

Note: Data for the period 1991-2012 was aggregated into two separate networks, each spanning 11 years. Presented in the table is the total number of directed weighted links for each separate year of the data, and for the two aggregate networks I and II. Data before 1991 is very sparse, with less than 1000 directed links per year during the period 1990-1981 and less than 100 before 1981.

Analysis. First, a dynamics-oriented approach is truer to the object under study. A citation is not a fixed structural feature such as a road in an infrastructure network. Citations represent knowledge flows. To be more precise, a citation represents transfer of a unit of knowledge per unit of time, from one inventor to another.9 Second, the dynamic modeling approach lets us incorporate two essential features of knowledge networks: link weighting and directionality. Flow-based analysis is particularly well adapted to the study of directed, weighted networks, such as trade (McNerney, Fath and Silverberg, 2013) and citation (Bohlin et al., 2014) networks, and has previously been applied to bibliometric knowledge networks (Bohlin et al., 2015; Lambiotte and Rosvall, 2012).

B. Knowledge Flow as a Measure of Domain Importance

A node’s importance to network dynamics is determined by its topological position within the network and by its intrinsic qualities. In social network analysis node importance has been defined in terms of centrality measures, such as degree centrality, closeness centrality, betweenness centrality, and other network-theoretic measures. Degree centrality is the number of links connected to a node. Proximity of a node to other nodes, defined as the inverse of the sum of shortest paths connecting the node

9A keen reader will note that much like patents themselves, citations are heterogeneous in their knowledge content. Although a metric for the quantity of information transfer at citation level would certainly have separate uses, it is unnecessary for the purpose of modeling knowledge flows between large aggregated entities, such as technology domains, where by operation of the law of large numbers total knowledge transfer will be approximated by the per-citation average times the total number of citations. Since the average per-citation knowledge coefficient can be normalized to 1, the maxim “1 citation = 1 unit of knowledge / unit of time” is a defensible assumption for the kind of aggregate entity flow analysis implemented in the present study.
to all other nodes on the network, is known as *closeness centrality*. *Betweenness centrality* of a node reflects the number of times it acts as a bridge between other nodes, or, more precisely, the number of inter-nodal shortest paths on the network that pass through it.

These common measures have a connection to flow dynamics on the network. The greater the degree of a node, the more flow can pass through it. Likewise, a central or bridging node can be a more capacious conduit of flow as a result of its topological position. The relationship between network flow and topological measures has been explored previously in Freeman, Borgatti and White (1991), Newman (2005) and Borgatti (2005), with the latter study concluding that “most commonly used centrality measures are not appropriate for most of the flows we are routinely interested in” (Ibid., p. 55).

Ultimately, it is the contribution of the node to network flow—by virtue of all its structural attributes—that represents its importance. Furthermore, nodal importance can also be influenced by intrinsic attributes, such as node size and the vitality of inventive activity within the technological field. Simulating network flow directly is a way to estimate the contribution of each node to flow on the network.

In the context of a knowledge network, a flow-based analytic approach lets us identify technologies that receive large volumes of knowledge from other technological fields. Modeling knowledge dynamics of the global innovation network provides an empirical way to estimate the learning intensity for the set of technological domains that constitute the network, which in turn gives us a way to rank technologies in terms of their importance to the global technology network.

**C. Knowledge Flow: From Heuristics to Simulation of the Stationary Distribution**

We can think of the random walker as a unit of knowledge circulating through the knowledge network. Below, we outline the heuristics for the random walk process that will simulate flow on the network. First, flux is greater on heavily weighted links than on lighter links. Thus, the greater the total strength of all in-links connected to a node, the more frequently the random walk will return to the node, the more time the random walker will spend on the node, and the greater will be the node’s flow estimate. Technologies that are knowledge-absorbing “learning hubs” will tend to cite more frequently.

Knowledge generated on the node also contributes to nodal flux. The more innovations in a technology domain cite other innovations from within the domain, the greater the importance of autocatalytic knowledge dynamics and the longer the random walker will remain on the node once it arrives.

Building on the above heuristics, we simulate knowledge flow on the global patent network with a random walker that traverses the network, stepping onto nodes with a frequency proportional to the weight of their directed links. The conditional probability with which a random walker moves from node \(i\) to node \(j\) is given by the weight of the directed link connecting \(i\) to \(j\), relative to the total weight of all out-links projecting from \(i\) to its neighbors:

\[
p_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}.
\]

Since self-links are included in the network structure, in the next step of the random walk the random walker can remain on node \(i\) with a probability proportional to the relative weight of the self-link.

Dynamics on the network can be measured with the steady-state distribution of the random walker on the network. With links drawn in the direction of knowledge flow, the random walker will tend to return to comparatively more knowledge-absorbing nodes that serve as learning hubs on the network. Changes in knowledge flow estimates will also document changes in intensity of innovative activity on the node.

If the network is irreducible and aperiodic, the flux generated by the random walker on the network will eventually ap-
proach a unique stationary distribution.\textsuperscript{10} However, if the network has disconnected or weakly connected components, the estimate of the steady-state distribution of the random walker on the network can be sensitive to the identity of the starting node. To overcome this problem, dynamics on complex networks can be simulated with a Markov process that includes a “teleportation parameter” $\tau$ which represents the probability with which the random walker jumps from node $i$ to any other node in the network, chosen at random. Teleportation allows the random walker to get “unstuck” from a weakly connected or disconnected node, ensuring a unique steady-state solution.\textsuperscript{11} 

\begin{equation}
    p_i = (1-\tau) \lambda_i + \tau \cdot h_i
\end{equation}

The stationary distribution of the random walker on the nodes of the network is given by the system of equations in Eq. (4), where the first term on the right-hand-side represents the probability that the random walker walks over to node $i$, while the second term represents the probability that the random walker arrives on node $i$ as a result of teleportation. Here, $1 - \tau$ is the fraction of time our random walker walks between nodes, while $\tau$ is the fraction of time he hops to a node. Terms $\lambda_i$ and $h_i$ are conditional probabilities of arriving on node $i$ by “walking” or through teleportation, respectively.

The probability the random walker steps onto node $i$ is given by the multiple of $p_j$ (the probability the walker is on node $j$) and $p_{ji}$ (the conditional probability with which the random walker moves from node $j$ to node $i$), summed over all $j$, as per Eq. (5):

\begin{equation}
    \lambda_i = \sum_j p_j \cdot p_{ji}.
\end{equation}

Probability $p_{ji}$ can be decomposed further, as function of the weight of the directed link from $j$ to $i$, relative to the total weight of all out-links projecting from $j$ to neighboring nodes (analogous to the heuristic in Eq. (3)):

\begin{equation}
    p_{ji} = \frac{w_{ji}}{\sum_k w_{jk}}.
\end{equation}

We now have all the elements in place for a complete description of the walking process and turn to set the rules governing teleportation.

Teleportation is more likely to nodes that play an important role on the network. Therefore, the conditional probability of teleporting to a node on the network can be set proportional to the sum of weights of the node’s out-links, as per Eq. (7):

\begin{equation}
    h_i = s^\text{out}_i.
\end{equation}

The value $s^\text{out}_i$ is also known as the out-strength of the node; it is defined as:

\begin{equation}
    s^\text{out}_i = \sum_j w_{ij}/\sum_{i,j} w_{ji}.
\end{equation}

The stationary distribution of knowledge across technology domains is given by the solution to the following system of equations:

\begin{equation}
    p_i = (1-\tau) \sum_j p_j \left( \frac{w_{ji}}{\sum_k w_{jk}} \right) + \tau \left( \sum_j w_{ij}/\sum_{i,j} w_{ji} \right)
\end{equation}

which is derived by substituting Eq. (6) into Eq. (5), Eq. (8) into Eq. (7) and Eqs. (5) and (7) into Eq. (4). In Eqs. (3)-(9) indices $i,j$ and $k$ are over the full range of knowledge categories in set $O$.

We solve the system in Eq. (9) via Von Mises iteration. The solution provides the steady-state distribution of knowledge flow across the nodes of the global technology network, which we describe and discuss in the next section.

\textsuperscript{10}A network is irreducible if there exists a path from every node to every other node. It is aperiodic if for the set of cycles on the network the greatest common denominator is 1.

\textsuperscript{11}See Rosvall and Bergstrom (2008) for additional discussion of teleportation.
VI. Results

In the core part of our analysis we simulate with a random walk the flow of ideas on the global innovation system. We take the empirical network as is, retaining its directed, weighted, reflexive aspects. Links are set in the direction of knowledge flow, that is, from the cited to the citing patent category. We keep self-links, which turn out to be an important factor in knowledge dynamics. We find that for technological domains knowledge sourced internally accounts for 43 percent of total link weight on the network during the 1991-2001 period, and 41 percent in the 2002-2012 period. Analysis of network topology excluding knowledge dynamics on the nodes would, therefore, leave out of view a large part of the complete picture.

Empirical simulation of knowledge dynamics on our networks requires specification of $\tau$. A teleportation rate at or near 0 can imply an absence of a stable stationary distribution. To estimate learning intensity of technology domains, $\tau$ must be low, yet above 0. For each empirical knowledge network we run the simulation under a number of alternative assumptions for $\tau$, including 0, 0.01, 0.05, 0.1, 0.15, 0.25, 0.35 and 0.5. While nodal flow estimates are sensitive to the teleportation parameter, their relative ranking changes very little as a result of changes in the teleportation probability.

As a robustness check, we estimate stationary distributions and derive corresponding rankings for the vector of technology domains, under a number of assumptions for the teleportation parameter. We then compare all sets of estimates to each other. Table 3 presents robustness checks for technology flow and rank estimates under alternative values of $\tau$. In each row, the assumed $\tau$, given in column (2), is compared to the baseline $\tau = 0.15$. Column (3) presents a summary measure of discrepancies between baseline and alternative vectors of flow estimates, in the form of the mean absolute difference of flow estimates divided by the mean baseline stationary distribution. The more dissimilar are the two vectors, the greater the value in column (3).

On average, the differences in flows calculated under alternative $\tau$ is no larger than 8 percent of the mean nodal flow estimate for Network I and no larger than 7 percent for Network II. Rankings based on flow estimates are compared in columns (4) and (5) of the table, which present Spearman and Kendall rank correlations, respectively. The closer the correlation coefficient is to 1, the more similar are the two rankings. We find that the importance ranking of technology domains is highly stable. In the remainder of the paper we focus on estimates obtained from simulations performed with the baseline $\tau = 0.15$.

A. Ranking Technological Domains by Knowledge Flow

Over time we can observe a slight shift in the distribution of knowledge flow across nodes of the network, towards greater concentration. Figure 2 presents a Lorenz-type curve showing the distribution of knowledge across technology domains. The $x$-axis represents the cumulative proportion of technology domains, sorted by knowledge flow, from lowest to highest. The cumulative share of knowledge flowing through the bottom $x$ domains is on the $y$ axis. Network I (dotted black line) is the cumulative distribution in the 1991-2001 period. Network II (solid red line) is the knowledge flow distribution in the 2002-2012 period. The slight shift of the Network II distribution inward indicates that less learning intensive technologies represented an even lower proportion of overall flow in the latter period. This shift is present throughout the distribution, but is most pronounced at the very top. The top 50 nodes of Network I contribute 66.3 percent of knowledge on the network; in Network II the top 50 subclasses represent 66.9 percent of circulated knowledge. During that time span a single subclass increased its contribution to overall network dynamics from 7.6 percent to 10.7 percent.

Tables 4 and 5 present the estimated knowledge flow on the global innovation network during the 1991-2001 and 2002-2012 periods, respectively. Technology domains are ranked according to their absorp-
Table 3—Flow and Rank Estimates Under Alternative Parameters $\tau$, Comparison to Baseline $\tau = 0.15$

<table>
<thead>
<tr>
<th>Network</th>
<th>$\tau$</th>
<th>Mean Absolute Difference in Distribution / Mean</th>
<th>Spearman Rank Correlation</th>
<th>Kendall Rank Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.00</td>
<td>0.082</td>
<td>0.999</td>
<td>0.978</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.043</td>
<td>1.000</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.018</td>
<td>1.000</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.024</td>
<td>1.000</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
<td>0.040</td>
<td>1.000</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.056</td>
<td>1.000</td>
<td>0.983</td>
</tr>
<tr>
<td>II</td>
<td>0.15</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.021</td>
<td>1.000</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
<td>0.035</td>
<td>1.000</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.051</td>
<td>0.999</td>
<td>0.982</td>
</tr>
</tbody>
</table>

Note: Insert note here.

Figure 2. Change in Distribution of Knowledge Across Technology Domains

Note: Presented is a Lorenz-type curve showing the distribution of knowledge flow across technology domains. The $x$ axis represents the cumulative proportion of technology domains, sorted by knowledge flow, from lowest to highest. The cumulative share of knowledge flowing onto the bottom $x$ domains is on the $y$ axis. Network 1 (dotted black line) is the cumulative knowledge flow distribution across technologies in the 1991-2001 period. Network 2 (solid red line) is the knowledge flow distribution in the 2002-2012 period.
tion of knowledge on the network, as measured by nodal flow. The first column of the two tables represents subclass rank. Column (2) represents change in subclass ranking between the two long decades: 1991-2001 ranking minus the rank in 2002-2012. By construction, a negative integer in column (2) represents a drop in subclass importance over the past two decades, while a positive integer indicates appreciation in ranking. The next two columns provide the IPC code and description for each subclass, based on the original IPC definitions.

To enable search, the IPC system contains elaborate definitions for each subclass, precisely delineating the types of technologies included and excluded. For macro-level scientometric purposes much of the detail is not essential. The subclass descriptions presented in column (4) are abridged, providing only a general indication of the technological content of each subclass.\(^{12}\)

Column (5) contains a measure of the flow of knowledge on the node—the focal indicator of this analysis—representing the intensity of innovation on each technological domain. This contribution is measured as the proportion of time the simulated random walker spends on the node in the steady-state, a subject discussed in the prior section. The proportions sum to 1 when all domains are taken into account.

The last two columns are based on the count of patents associated with each IPC subclass. Column (6) presents the patents allied to each subclass as a proportion of all patents granted during the period—a simple measure of relative subclass size. Column (7) contains the total number of patents granted in each subclass. Columns (6) and (7) allow us to evaluate the extent to which knowledge flows, dissociated from intrinsic technology attributes and network position, are a measure of the importance of patent categories to innovation dynamics. In many cases, subclasses that do not rank highly on patent-count-based measures of importance, turn out to be vital to the knowledge network (for example, see subclass A61F). In other words, the measure of knowledge flow adds new information not present in columns (6) and (7) of Tables 4 and 5.

B. Identifying “Learning Hubs”

Tables 4 and 5 give an overview of the evolution of technology, measured by shifts in the contribution of technological domains to the circulation of knowledge on the global innovation network. Technological trends during the recent decades display a mix of change and continuity. At the core of the global innovation system lie technological specialties that contribute significantly to technological progress, and the importance of which persists over a long period of time. The contribution of the more peripheral technological domains is more fluid, with some domains undergoing large shifts in importance over time.

Drawing a line between “learning hubs” and all others is somewhat arbitrary if the placement is based solely on the ranking. However, if we add to the concept some measure of persistence, the list is more straightforward to define. The top 7 subclasses in Tables 4 and 5 consistently play an oversized role in knowledge dynamics on the innovation network. These 7 learning-hub technologies each constitute between 2 and 11 percent of flow on the network. Further, going from the 1991-2001 to the 2002-2012 network, we see that the list of the top 7 subclasses has remained remarkably stable.

What are the most important technologies of our time? As this analysis shows, if importance is defined in terms of contribution to general technological progress as mediated by cross-disciplinary learning patterns, in both decades, among the top 7 subclasses one finds select technologies focusing on information, computing and telecommunication (“Electric digital data processing” [G06F], “Semiconductor and electric solid state devices” [H01L], “Transmission of digital information” [H04L], “Pictorial communication” [H04N]) and categories dealing with medicine ("Diag-
Table 4—Network 1 (1991-2001), Ranking of Subclasses by Knowledge Flow

<table>
<thead>
<tr>
<th>Rank</th>
<th>Δ IPC</th>
<th>Subclass</th>
<th>Abridged Description</th>
<th>Flow (Prop.)</th>
<th>Total Patents (1,000s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>G06F</td>
<td>Electric digital data processing</td>
<td>0.076</td>
<td>0.031</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>H01L</td>
<td>Semiconductor devices; electric solid state devices</td>
<td>0.050</td>
<td>0.039</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>A61B</td>
<td>Diagnosis; surgery; identification</td>
<td>0.043</td>
<td>0.013</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>A61K</td>
<td>Preparations for medical, dental, or toilet purposes</td>
<td>0.041</td>
<td>0.029</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>H04L</td>
<td>Transmission of digital information, e.g., Telegraphy</td>
<td>0.035</td>
<td>0.015</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>A61F</td>
<td>Filters implantable into blood vessels; prostheses</td>
<td>0.030</td>
<td>0.009</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>H04N</td>
<td>Pictorial communication, e.g., Television</td>
<td>0.023</td>
<td>0.022</td>
</tr>
<tr>
<td>8</td>
<td>-2</td>
<td>G01N</td>
<td>Determining chemical or physical properties of materials</td>
<td>0.022</td>
<td>0.017</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>G06Q</td>
<td>Data processing, esp. administrative and financial</td>
<td>0.020</td>
<td>0.003</td>
</tr>
<tr>
<td>10</td>
<td>-1</td>
<td>A61M</td>
<td>Devices for introducing media into, or onto, the body</td>
<td>0.019</td>
<td>0.007</td>
</tr>
<tr>
<td>11</td>
<td>-7</td>
<td>G11B</td>
<td>Information storage (based on movement)</td>
<td>0.016</td>
<td>0.017</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>G02B</td>
<td>Optical elements, systems, or apparatus</td>
<td>0.015</td>
<td>0.012</td>
</tr>
<tr>
<td>13</td>
<td>-2</td>
<td>C07D</td>
<td>Heterocyclic compounds</td>
<td>0.012</td>
<td>0.013</td>
</tr>
<tr>
<td>14</td>
<td>-14</td>
<td>C12N</td>
<td>Micro-organisms or enzymes; compositions thereof</td>
<td>0.011</td>
<td>0.008</td>
</tr>
<tr>
<td>15</td>
<td>-1</td>
<td>H04B</td>
<td>Transmission of information</td>
<td>0.011</td>
<td>0.010</td>
</tr>
<tr>
<td>16</td>
<td>-3</td>
<td>H04M</td>
<td>Telephonic communication</td>
<td>0.011</td>
<td>0.007</td>
</tr>
<tr>
<td>17</td>
<td>-7</td>
<td>A61P</td>
<td>Therapeutic chemical compounds or preparations</td>
<td>0.010</td>
<td>0.011</td>
</tr>
<tr>
<td>18</td>
<td>-4</td>
<td>B41J</td>
<td>Typewriters; selective printing mechanisms</td>
<td>0.010</td>
<td>0.008</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>A61N</td>
<td>Electro-, magneto-, radiation- and ultrasound-therapy</td>
<td>0.010</td>
<td>0.003</td>
</tr>
<tr>
<td>20</td>
<td>-7</td>
<td>B65D</td>
<td>Containers for storage or transport</td>
<td>0.009</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Note: Network 1 represents all patents granted during the 1991-2001 period, aggregated into 4-digit patent subclasses (nodes), along with inventor-origin citations between subclasses (links). The empirical knowledge network consists of 629 subclasses. Presented are the top 50 subclasses as measured by knowledge flow. Knowledge flow on the network is simulated with random walks. Column descriptions: (1) Subclass rank in Network 1 (1991-2001). (2) Change in rank: rank in 1991-2001 minus rank in 2002-2012. (3) IPC Version 8 subclass code. (4) Abridged description, based on IPC Version 8 definition, but abridged for compactness. (5) Proportion of network knowledge flowing onto the node representing the subclass. (6) Subclass patents as proportion of all patents granted during the 1991-2001 period. (7) Number of patents granted during the 1991-2001 period assigned to subclass, in thousands (fractional count). An extended version of this table containing all subclasses is available from the author.
Table 5—Network 2 (2002-2012), Ranking of Subclasses by Knowledge Flow

<table>
<thead>
<tr>
<th>Rank</th>
<th>IPC Subclass</th>
<th>Abridged Description</th>
<th>Flow (Prop.)</th>
<th>Total Patents (1,000s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G06F</td>
<td>Electric digital data processing</td>
<td>0.107</td>
<td>0.051</td>
</tr>
<tr>
<td>2</td>
<td>H01L</td>
<td>Semiconductor devices; electric solid state devices</td>
<td>0.059</td>
<td>0.052</td>
</tr>
<tr>
<td>3</td>
<td>A61B</td>
<td>Diagnosis; surgery; identification</td>
<td>0.043</td>
<td>0.015</td>
</tr>
<tr>
<td>4</td>
<td>A61K</td>
<td>Preparations for medical, dental, or toilet purposes</td>
<td>0.031</td>
<td>0.024</td>
</tr>
<tr>
<td>5</td>
<td>H04L</td>
<td>Transmission of digital information, e.g. Telephony</td>
<td>0.026</td>
<td>0.025</td>
</tr>
<tr>
<td>6</td>
<td>A61F</td>
<td>Filters implantable into blood vessels; prostheses</td>
<td>0.023</td>
<td>0.007</td>
</tr>
<tr>
<td>7</td>
<td>H04N</td>
<td>Pictorial communication, e.g. Television</td>
<td>0.018</td>
<td>0.023</td>
</tr>
<tr>
<td>8</td>
<td>G06K</td>
<td>Recognition and presentation of data</td>
<td>0.017</td>
<td>0.008</td>
</tr>
<tr>
<td>9</td>
<td>G06Q</td>
<td>Data processing, esp. administrative and financial</td>
<td>0.017</td>
<td>0.007</td>
</tr>
<tr>
<td>10</td>
<td>G01N</td>
<td>Determining chemical or physical properties of materials</td>
<td>0.016</td>
<td>0.018</td>
</tr>
<tr>
<td>11</td>
<td>G01M</td>
<td>Devices for introducing media into, or onto, the body</td>
<td>0.016</td>
<td>0.006</td>
</tr>
<tr>
<td>12</td>
<td>G02B</td>
<td>Optical elements, systems, or apparatus</td>
<td>0.015</td>
<td>0.013</td>
</tr>
<tr>
<td>13</td>
<td>G11C</td>
<td>Information storage (static)</td>
<td>0.014</td>
<td>0.008</td>
</tr>
<tr>
<td>14</td>
<td>E21B</td>
<td>Earth or rock drilling</td>
<td>0.013</td>
<td>0.006</td>
</tr>
<tr>
<td>15</td>
<td>C07D</td>
<td>Heterocyclic compounds</td>
<td>0.013</td>
<td>0.010</td>
</tr>
<tr>
<td>16</td>
<td>H04B</td>
<td>Transmission of information</td>
<td>0.012</td>
<td>0.014</td>
</tr>
<tr>
<td>17</td>
<td>A61N</td>
<td>Electro-, magneto-, radiation- and ultrasound-therapy</td>
<td>0.011</td>
<td>0.003</td>
</tr>
<tr>
<td>18</td>
<td>G11B</td>
<td>Information storage (based on movement)</td>
<td>0.011</td>
<td>0.013</td>
</tr>
<tr>
<td>19</td>
<td>H04M</td>
<td>Telephonic communication</td>
<td>0.010</td>
<td>0.007</td>
</tr>
<tr>
<td>20</td>
<td>H04W</td>
<td>Wireless communication networks</td>
<td>0.010</td>
<td>0.011</td>
</tr>
</tbody>
</table>

Note: Network 2 represents all patents granted during the 2002-2012 period, aggregated into 4-digit patent subclasses (nodes), along with inventor-origin citations between subclasses (links). The empirical knowledge network consists of 637 subclasses. Presented are the top 50 subclasses as measured by knowledge flow. Knowledge flow on the network is simulated with random walks. Column descriptions: (1) Subclass rank in Network 1 (2002-2012). (2) Change in rank: rank in 1991-2001 minus rank in 2002-2012. (3) IPC Version 8 subclass code. (4) Subclass description, based on IPC Version 8 definition, but abridged for compactness. (5) Proportion of network knowledge flowing onto the node representing the subclass. (6) Subclass patents as proportion of all patents granted during the 2002-2012 period. (7) Number of patents granted during the 2002-2012 period assigned to subclass, in thousands (fractional count). An extended version of this table containing all subclasses is available from the author.
nosis, surgery and identification" [A61B], “Preparations for medical, dental or toilet purposes” [A61K], “Filters implantable into blood vessels and prostheses” [A61F]). Altogether, the above 7 subclasses account for 30.7 (29.8) percent of all knowledge flow, and 19.7 (15.8) percent of all patents granted in the 2002-2012 (1991-2001) period.

Although the importance of information and telecommunication technology (ICT) is widely recognized, we see from our results that only specific technological domains within the broader ICT constellation of technologies have a key role in the global innovation system. For example the subclass “Multiplex communication” [H04J], also an example of an ICT technology, contributes to the global innovation system less than \( \frac{1}{4} \)th the knowledge flow of “Pictorial communication” [H04N]. The contribution to network knowledge flow of “Optical computing devices” [G06E] and “Plasma technique” [H05H] is even more modest, as these subclasses do not even appear on the top 50 list. These technological domains might still have an important role to play locally, within their ego network or within the functional module of which they are a part, but individually, they have limited influence on the network as a whole.

VII. Discussion and Conclusion

This study started with an overarching question: what technologies are the most important for technological progress? The answer is certainly of intellectual interest to economic historians and scholars of science and technology. It is also of practical consequence for technology management and innovation policy.

Technology managers require information about past development, current position, and likely future trajectory of technologies in their portfolio. By synthesizing techniques from patentometrics, dynamical network models, and probability theory, this study went beyond conventional methods, such as simple or composite indexes used in monitoring technology trends. One of the contributions of the presented methodology is its ability to take into account the interactive and relational nature of innovative activity. Since technologies do not develop in isolation, but source knowledge from, and contribute to, other technologies, the network dynamics methodology provides a way to study technological domains in their context.

To answer the guiding question we model technological knowledge as a dynamic network, which allows identification of technology domains that are the global center of absorption and recombination of technological knowledge, for which we adopt the term “learning hubs.” The sphere of technologically relevant knowledge is conceptualized as a reflexive, directed, link- and node-weighted complex network, with distinct spheres of knowledge (or technology domains) representing network nodes and learning (or knowledge flows) across domains acting as inter-nodal links.

The empirical knowledge network in this study is constructed from a sweeping patent database. We instantiate the nodes of the knowledge network from patent categories of the IPC. For the purpose of capturing information on knowledge flows, we isolate the citations that are most closely associated with inventor knowledge acquisition. Links between technology domains, representing knowledge transfer between fields of technology, are constructed from patent citations provided by inventors, aggregated at the patent subclass level. Inventor-origin citations are so far the best proxy for knowledge flow on patent networks validated by prior empirical studies. We also de-duplicate the patent record, extract unique inventions, and hone in on inventor knowledge at the time of invention.

Modeling the evolving dynamics of knowledge flow on the world knowledge network allows us to identify technologies that are making the greatest contribution to dynamics of technological progress and reveals technology trends of the past two decades. Furthermore, by comparing knowledge dynamics along the network during different time periods we extract information about the evolution of the knowledge network.

Among the 7 learning hub technologies,
our analysis identifies four technological categories that can be broadly described under the term “ICT,” and three medical technology subclasses. Among other transformations in the global system of innovation, the past two decades have witnessed the relative waning in the importance to the global innovation system of medical technologies, while ICT learning hub technologies have increased in importance. Although medical learning hub technologies remain important, they have lost some ground in the global innovation system.

While our measure of knowledge flows on nodes of the global innovation network captures in a summary way all information about the relational structure and dynamics of the network, there exist a number of questions we have barely touched, that could also be answered by application of similar methods. For example focus on dynamics could also reveal interaction between specific components of the network. We place identification and exploration of modular sub-components of the global knowledge network on our future research agenda and in a forthcoming companion study, we explore the evolving community structure of the global innovation network.

REFERENCES


Bohlin, Ludvig, Alcides Viamontes Esquivel, Andrea Lancichinetti, and Martin Rosvall. 2015. “Robustness of journal rankings by network flows with different amounts of memory.” *Journal of the Association for Information Science and Technology*.

Bohlin, Ludvig, Daniel Edler, Andrea Lancichinetti, and Martin Rosvall. 2014. “Community Detection and Visualization of Networks with the Map Equation Framework.”


Podolny, Joel M, Toby E Stuart, and Michael T Hannan. 1996. “Networks, knowledge, and niches: Competition in


