Overlay technology space map for analyzing design knowledge base of a technology domain: the case of hybrid electric vehicles

Binyang Song1 · Bowen Yan1,2 · Giorgio Triulzi2,3 · Jeffrey Alstott2,4 · Jianxi Luo1,2

Received: 17 October 2016 / Revised: 6 September 2018 / Accepted: 26 February 2019 / Published online: 7 March 2019
© Springer-Verlag London Ltd., part of Springer Nature 2019

Abstract
A tangible understanding of the latent design knowledge base of a technology domain, i.e., the set of technologies and related design knowledge used to solve the specific problems of a domain, and how it evolves, can guide engineering design efforts in that domain. However, methods for extracting, analyzing and understanding the structure and evolutionary trajectories of a domain’s accumulated design knowledge base are still underdeveloped. This study introduces a network-based methodology for visualizing and analyzing the structure and expansion trajectories of the design knowledge base of a given technology domain. The methodology is centered on overlaying the total technology space, represented as a network of all known technologies based on patent data, with the specific knowledge positions and estimated expansion paths of a specific domain as a subgraph of the total network. We demonstrate the methodology via a case study of hybrid electric vehicles. The methodology may help designers understand the technology evolution trajectories of their domain and suggest next design opportunities or directions.

Keywords Design theory · Data-driven design · Knowledge management · Technology evolution · Network analysis and visualization · Hybrid electric vehicles

1 Introduction
Technology domains, such as semiconductors, robotics and hybrid electric vehicles (HEVs), encompass various technologies that address interconnected engineering design problems and perform closely related functions (Layton 1974; Dosi 1982; Yayavaram and Ahuja 2008; Dibiaggio et al. 2014). As the engineers in a domain continually explore and synthesize additional technologies beyond the domain to create new solutions to solve the problems of their domain (Arthur 2009), the scope of engineering design knowledge in the domain expands. For instance, the design of motor vehicles, initially invented in the late nineteenth century, was originally based solely on machine elements but has incrementally synthesized composite materials, electronics, communication, signaling and computing technologies over the course of its evolution. For engineers who specialize in a particular domain, it is beneficial but difficult to understand the structure and trajectories of the technological expansion of their domains and to identify potential technologies that will be useful for future design and innovation.

To address this challenge, this paper introduces a data-driven methodology for estimating, visualizing and assessing a domain’s design knowledge base. Herein, a design knowledge base is the set of technologies and related design knowledge used to solve the specific problems of the domain. In particular, the design theory has explicated the impact of the prior knowledge base on next designs (Hatchuel and Weil 2009), and prior empirical studies have shown that technologies based on more related knowledge bases are more likely to be synthesized to yield inventions (Fleming 2001; Yayavaram and Ahuja 2008; Dibiaggio et al. 2014; He and Luo 2017). Following these prior understandings, our methodology is focused on the relatedness between
technologies for the analyzing the design knowledge base of a domain.

Specifically, the methodology is centered on overlaying the total technology space, represented as a network of all known technologies, with the design knowledge base of a specific domain as a subgraph of the total network. In such a network, technologies are operationalized as patent classes and linked according to their pairwise relatedness that can be measured using different metrics based on patent data. To address the uncertainty in selecting a relatedness metric, we introduce a statistical technique to identify the best relatedness metric and the resulting network with the highest explanatory power on the historical expansions of a domain’s design knowledge base. Then, the resulting network is used as a total technology space map to estimate, visualize and analyze the structure and evolution trajectories of the domain’s design knowledge base as a subgraph of its.

Such an overlay network provides a macro and tangible picture of the relative positions and expansion trajectories of a domain’s design knowledge base, and the various unexplored technologies beyond the design knowledge base, as well as their relatedness to the design knowledge base. It enables graph theory-based network analysis and data-driven visualizations that can help designers and managers better understand the structure and evolution trajectories of their domains or firms in the domain, compare expansion characteristics of different domains and firms, and suggest potential unexplored technologies for next design opportunities. We demonstrate the methodology and show its utilities via a case study of the HEV domain and the leading firms in the domain.

2 Related work

2.1 Design knowledge expansion

A large body of literature has suggested that new designs often arise via the analogical transfer, recombination, blending or synthesis of existing technologies and their related knowledge (Fleming 2001; Fleming and Sorenson 2001, 2004; Weisberg 2006; Arthur 2009; Taura and Nagai 2012; Tang and Luo 2013). Meanwhile, such analogy, combination or synthesis is often made across different technical fields (Fleming 2001; Fleming and Sorenson 2001; Yayavaram and Ahuja 2008). In engineering design, researchers have proposed a few methodologies for facilitating design processes across fields or disciplines. For instance, infused design brings design problem representation up to a mathematical meta-level to facilitate the use of knowledge, methods and solutions across various technological fields (Shai and Reich 2004). Reich and Shai (2012) also developed the concept of an interdisciplinary engineering knowledge genome to aid in the retrieval of common knowledge and method structures in different technological fields. Likewise, design-by-analogy draws existing solutions and related design knowledge from a source field to solve design problems in a target field (Linsey 2007; Linsey et al. 2012).

As the additional technological knowledge is analogized, synthesized or transferred to generate new solutions and thus new design knowledge in a domain, the domain’s design knowledge base may expand across technologies in the total technology space (Baxter et al. 2007; Zdrahal et al. 2007; Reich and Shai 2012). Such an expansion of the design knowledge base of a domain resonates with the expansion of the K-space (where K denotes knowledge) during a design process in the C–K theory (Hatchuel and Weil 2003, 2009; Hatchuel et al. 2004). The design knowledge space expansion process raises a question regarding whether and how the structure of the prior design knowledge base underlying the design process might influence the prospects and directions of its expansion in the total technology space.

Note that our reasoning and analysis herein primarily focus on design knowledge rather than scientific knowledge. Luo (2015) argued that the design process may use both types of knowledge as inputs. Design knowledge is related to the existing solutions, methods, tools and practices and any other artifacts in a domain of interest, whereas scientific knowledge involves the understanding of the natural phenomena related to the design problems of the domain. For example, the design of solar photovoltaic (PV) cells requires extensive domain-specific design knowledge regarding existing solar PV cell designs, manufacturing methods, tools, and processes in addition to the basic scientific knowledge regarding natural phenomena of the photoelectric effect. Design knowledge is generated and accumulated during the design process, as suggested by the C–K theory (Hatchuel and Weil 2009), whereas scientific knowledge is typically created during the scientific discovery process, which is driven by curiosity.

2.2 Influence of knowledge relatedness

Technologies are not isolated but related to each other to different degrees. In turn, the relatedness between technologies may condition or enable the prospects of their analogies, synthesizes or combinations, and the expansion of a domain’s design knowledge base. Meanwhile, the literature has suggested that relatedness between various technologies is a matter of degree and takes various forms, such as interdependency, complementarity and similarity (Jaffe 1986; Fleming 2001; Fleming and Sorenson 2001; Menamee 2013; Dibiaggio et al. 2014; Alstott et al. 2017b).

For instance, Fleming and Sorenson (2001) analyzed the ease of recombination of knowledge components as the reverse proxy of their “interdependency” and found that an
intermediate degree of interdependency benefits the usefulness of inventions. Yayavaram and Ahuja (2008) investigated the decomposability of a firm’s knowledge base using a technology network and evidenced that a nearly decomposable knowledge base increases inventions’ usefulness and the knowledge base’s adaptability with respect to change. Uzzi et al. (2013) found that the most impactful scientific knowledge was produced primarily based on conventional combinations (measured as the historical combination frequency) of prior knowledge. Dibiaggio et al. (2014) studied the complementarity and substitutability of a firm’s knowledge elements and found that complementarity positively contributes to invention quality, whereas substitutability is beneficial to explorative invention.

Prior studies of design creativity have further suggested that it is easier and thus more likely for a designer to understand and successfully use new knowledge that is more related or similar to what he or she has already mastered (Wuyts et al. 2005; Chan and Schunn 2015; Srinivasan et al. 2017). For instance, Srinivasan et al. (2017), via a concept generation experiment, found that designers preferred to utilize external patent stimuli proximate to rather than distant from the design problem. Fu et al. (2013b) constructed Bayesian networks of patents based on functional similarities and found that patents “near” the design problem in the network are more effective than those “far away” in stimulating new design solution concepts. Via an analysis of thousands of design concepts from a web-based open innovation challenge platform, Chan and Schunn (2015) found that sources of external design inspiration are associated with the novelty of generated ideas and conceptually closer sources of design inspiration, rather than sources that are farther away, are more beneficial to the design outcome. However, these findings do not deny the potential of combining distant pieces of knowledge for more novel designs (He and Luo 2017), though such potential is difficult to transform into successful outcomes (Fu et al. 2013b).

Researchers have also studied the influences of knowledge relatedness on the exploration and diversification of design agents at more macro levels. At the inventor level, Alstott et al. (2017a) analyzed the records of more than two million inventors in the patent database of the United States Patent and Trademark Office (USPTO) and found inventors are much more likely to enter new technology domains that are more related to their prior ones, and their relatedness is based on empirical direct citations between patent categories normalized against random citations. At the level of firms, Breschi et al. (2003) found European firms preferred to diversify into technologies that are more related to their core technologies, using co-classification similarity between technologies as proxy of knowledge relatedness. At the city level, Boschma et al. (2014) found that American cities mostly entered new technological areas that are more related to those they have entered, and their relatedness was calculated as the possibility of cities having more than average percent of patents in both technologies. Rigby (2015) measured the proximity between technologies based on citation data and concluded with similar findings.

In sum, prior studies have taken different perspectives to define and measure relatedness between technologies and implied the influence of knowledge relatedness on the evolution of design knowledge bases, i.e., the design knowledge base of a domain is more likely to expand to include new technologies that are more related to those already covered in that design knowledge base.

### 2.3 Relatedness-based patent class networks

Recently, patent data have been used to calculate knowledge relatedness among technologies and then construct large-scale network maps that represent the total technology space (Kay et al. 2014; Alstott et al. 2017b; Yan and Luo 2017a). In such network maps, nodes represent different technological fields or types of technologies, operationalized by patent classes defined by some official patent classification system, such as the International Patent Classification (IPC) and the Cooperative Patent Classification (CPC), and links represent the knowledge relatedness between patent classes. The total technology space map can be overlaid with the patent records of a design agent, i.e., an inventor or firm as done in Yan and Luo (2017b) and Luo et al. (2017), or the patent records related to a domain, as done in Kay et al. (2014) and Leydesdorff et al. (2014). The overlay technique enables the analysis of a domain’s design knowledge base as a subspace of the total technology space and the visual inspection of the expansion of the subspace within the total space.

The technology space maps in previous studies are based on a variety of knowledge relatedness metrics to quantify link weights between nodes in a network. One group of metrics uses patent reference information. For example, the reference-based Jaccard index is calculated as the number of shared references normalized by the number of all unique references of patents in either technology class (Jaccard 1901; Small 1973; Yan and Luo 2017a) and implies the overlapping of the knowledge bases of patent classes. The reference-based cosine similarity index measures the vector similarity of two technology classes’ distributions of citations in all patent classes (Jaffe 1986; Kay et al. 2014; Leydesdorff et al. 2014). Another group of metrics mines patent classification information. For instance, the classification-based cosine similarity index measures the vector similarity of two technology classes’ distributions of shared patents in all patent classes (Breschi et al. 2003; Ejermo 2015; Kogler et al. 2013). The normalized co-classification index measures the deviation of the empirically observed number of patents classified simultaneously in two technology classes.
Choose a technology space map with high explanatory power on the historical expansion of the domain’s design knowledge base

Overlay visualization of the technology space map with design knowledge base and expansion paths of the domain

Quantitative analysis on the structure and evolution of the design knowledge base of the domain

Identify potential technologies for next move of the domain

**Fig. 1** Flow diagram of our methodology

from the expected value that corresponds to a random classification (Teece et al. 1994; Dibiaggio and Nesta 2005). Interested readers may refer to a recent review and comparison of the most popular knowledge relatedness measures used in patent mapping (Yan and Luo 2017a). In this research, a key task is to identify the most suitable measure of knowledge relatedness among the available choices for the overlay mapping analysis of a specific domain of interest.

Taken together, prior studies have suggested the influence of knowledge relatedness on the expansion of design knowledge bases and the existence of various knowledge relatedness metrics for constructing technology network maps using patent data. On this basis, to further develop a network-based methodology for visualizing and analyzing a design knowledge base and its evolution, our research explores answers to the following questions: (1) which relatedness measure should be chosen for patent technology network mapping; (2) how to visualize a design knowledge base and estimate its expansion paths in the network map; (3) how can network metrics deliver insights on the structure and evolution of a domain’s design knowledge base.

### 3 Methodology

Specifically, our methodology involves the use of patent data to map the total technology space as a network of patent technology classes with a properly chosen knowledge relatedness metric, and to estimate, visualize and analyze the structure and expansion paths of a domain’s design knowledge base as a subgraph of the total network. Figure 1 depicts the major steps of our methodology regarding data retrieval, metric selection, mapping, visualization, quantitative and qualitative analysis.

Our methodology, based on patent data and network analysis, is also in line with the growing research involving the exploration of patent databases for data-driven design aids (Szykman et al. 2000; Mukherjea et al. 2005; Fu et al. 2013a). Differing from these prior studies that typically focused on small patent samples related to a design problem or a single patented technology, our method analyzes a technology domain and considers all technologies in the total technology space based on the entire USPTO patent database.

**STEP 1: Retrieve the patent data of a domain** For data-driven analyses of the design knowledge base of a technology domain, the first step is to retrieve the set of utility patents relevant to the design artifacts in the domain. Various techniques for retrieving the patents of a specific design domain have been proposed in the literature. The most intuitive approach is to search by keywords, which are often chosen based on the field knowledge or intuition of the researchers (Koch et al. 2009; Alberts et al. 2011). Koch et al. (2009) created a tool called PatViz that enables patent searchers to build complex search terms iteratively and visually. Boolean operators are widely used to build search queries by combining relevant keywords or excluding irrelevant keywords (Alberts et al. 2011). Semantic analysis has also been applied to patent retrieval (D’hondt 2009; Mukherjea et al. 2005). Other researchers have retrieved patents relevant to a design domain of interest by first identifying the notable firms or inventors in the domain and then searching the patents granted to them (Yayavaram and Ahuja 2008; Nakamura et al. 2015). Song and Luo (2017) recently proposed a heuristic and iterative method for retrieving patents relevant to a specialized technology domain by integrating the mining of patent texts, citation relationships and inventor information. Given the variety of patent retrieval methods in the literature, one needs to select a suitable method according to the characteristics of the domain and the specific needs of his or her analysis.

**STEP 2: Choose a technology space map** We use the entire USPTO patent database from 1976 to 2016 (July) to empirically create a technology network map that approximates and represents the total space of known technologies of human society to date. The data set contains 6,346,426 US utility patents, and each of them is classified into one or multiple IPC classes. In our patent technology network, nodes represent 3-digit IPC patent classes. IPC patent classes are defined at different aggregation levels, e.g., 3-digit level and 4-digit level. We chose 3-digit classes as nodes in the network to obtain the best resolution of the technology space and to achieve simple visualization and analysis. As a result, the network contains 122 nodes, i.e., 3-digit IPC classes, except for several undefined or ambiguous ones.
To construct the technology network, a measure of knowledge relatedness between technologies is required to quantify link weights. In the literature, various knowledge relatedness metrics have been proposed for different meanings (see Sect. 2.3). The best metric of knowledge relatedness might differ for different domains as the key drivers of technology development may vary in different contexts. For example, the reference-based Jaccard index measures the extent to which the knowledge bases of two technologies overlap; the reference-based cosine index measures the similarity of knowledge base distributions of different technologies; the classification-based cosine index measures substitutability of technologies; the normalized co-classification index measures complementarity. Therefore, the knowledge relatedness metric needs to be selected in a case-dependent manner. For this purpose, we propose a statistical technique to compare the alternative metrics in terms of the explanatory power of the resulting networks regarding the historical expansions of the focal domain’s design knowledge base.

Specifically, we investigate the likelihood that the domain’s design knowledge base has expanded to cover unexplored technologies with higher relatedness to it instead of other less related unexplored technologies. To calculate the likelihood, the relatedness percentile of each historical technology entry to the design knowledge base is calculated. Two steps are required to calculate the relatedness percentile:

1. identify the strongest link (i.e., the link with the highest relatedness) between a specific technology entry A and the previously covered technologies, e.g., B–A;
2. calculate A’s relatedness percentile as the rank percentile of the relatedness value of link B–A among the relatedness values of all links connecting the previously covered technology B and all unexplored technologies. The relatedness percentile indicates the percentage of links between technology B and any unexplored technologies that have a relatedness value equal to or lower than the relatedness value of link B–A.

Alternatively, the relatedness percentile can be calculated using the weighted average relatedness, i.e., the average relatedness value of an unexplored technology’s network links to all technologies previously covered in the design knowledge base, weighted by the number of domain-related patents for each covered technology. Thus, each unexplored technology will be characterized by such a weighted average relatedness percentile at the time it was first covered by the given domain’s design knowledge base.

Then, we compare the explanatory power of alternative network maps (resulting from alternative relatedness measures) by evaluating the accumulative distributions of the domain’s historical technology entries according to their relatedness percentiles calculated based on alternative knowledge relatedness measures. If a domain’s design knowledge base preferentially expands to cover the new technologies that are more related to those technologies already covered by the design knowledge base, the distribution of that domain’s historical new technology entries will be skewed toward a relatedness percentile of 100%. In contrast, if new technologies are explored randomly regardless of their relatedness to the previously established design knowledge base of the domain, the historical new technology entries will follow a uniform distribution. One can conduct a Kolmogorov–Smirnov test to check whether an observed distribution presents a statistically significant deviation from the reference uniform distribution.

The relatedness measure and the resulting network map that survives the test and yields the most skewed distribution have the highest explanatory power regarding the domain’s historical technology entries. They are further suggested for the analysis and visualization of the evolution trajectories of the domain. Note that for different technology domains and contexts, the most explanatory measure might differ and thus must be selected and validated on a case-by-case basis.

**STEP 3: Overlay the technology space map with the design knowledge base and its expansion trajectories of the domain for visualization**

To visualize and analyze the domain’s design knowledge base, we overlay the total technology space network selected in the previous step with the domain’s design knowledge base and its estimated expansion trajectories. On the background map, we first highlight the technology classes to which the patents of the domain are assigned. These highlighted classes represent the technologies constituting the design knowledge base of the domain. We further identify and highlight the most likely expansion paths from the technologies that were covered in the design knowledge base earlier to those covered later. Specifically, for each new technology entry, we highlight its strongest link to any previously covered technologies in the design knowledge base. These strongest links together suggest the most likely paths through which the design knowledge base expanded from earlier-covered to later covered technologies.

**STEP 4: Quantitative network analysis of the structure and evolution of the design knowledge base**

Once the domain’s design knowledge base has been located in the technology space map, graph-theoretic metrics can be used to assess the multi-node structure and positions of the design knowledge base within the network map. In the following, we introduce three such metrics as examples. Additional metrics can be developed for other purposes.

a. **Entropy**

The entropy of a design knowledge base is calculated using the information entropy formula (Allaire et al. 2012).
Entropy = \frac{1}{C} \sum_{i} x_i \ln \frac{1}{x_i}, \quad (1)

where \(x_i\) is the portion of domain-related patents in class \(i\) at a given time. \(C\) is a normalizing constant equal to the maximum attainable entropy corresponding to the configuration in which patents from domain \(z\) would be equally distributed across all technology classes \(i\). Entropy was initially a thermodynamics concept that is used to measure the thermodynamic stage of a system, with equilibrium corresponding to maximum entropy. It reveals to what extent the design knowledge base is diversified. High levels of entropy, close to the maximum value of 1, imply an increasing limitation of the design knowledge base for further movement and evolution.

b. Coherence

Coherence is calculated as the weighted average relatedness between technologies covered by a design knowledge base at a given point in time.

\[
\text{Coherence} = \sum_{i} \sum_{j(i \neq j)} x_i x_j \varphi_{ij} / \sum_{i} \sum_{j} x_i x_j, \quad (2)
\]

where \(\varphi_{ij}\) is the relatedness between technology classes \(i\) and \(j\). \(x_i\) is the portion of the domain-related patents in class \(i\) relative to all domain-related patents. According to the literature (e.g., Teece et al. 1994; Leten et al. 2007), we refer to this measure as the “coherence” among different technologies. A high coherence of the design knowledge base indicates that the designs in the corresponding domain are based on a set of technologies with high knowledge relatedness to each other and can be easily combined and recombined to yield new designs.

c. Expandability

Expandability is calculated as the weighted sum of the relatedness between the technologies within a domain’s design knowledge base and the remaining unexplored technologies in the technology space. We also normalize the “expandability” by its maximum attainable value, ensuring that the metric is in the range of [0, 1] and more comparable.

\[
\text{Expandability} = \sum_{i} \sum_{j(i \neq j)} x_i \varphi_{ij} s_j / A, \quad (3)
\]

where \(\varphi_{ij}\) is the relatedness between technology classes \(i\) and \(j\). \(x_i\) is the portion of the domain-related patents in class \(i\) relative to all domain-related patents; \(s_j\) is 1 if the domain has no patent in technology class \(j\) and is 0 otherwise, so that \(\varphi_{ij}\) is counted only when technology class \(i\) is in the domain and technology class \(j\) is unexplored. \(A\) is a normalizing constant and equal to the sum of the relatedness between the most connected technology class, i.e., the technology class with the highest weighted degree centrality in the network, and the other technology classes. Given the positions and structure of a design knowledge base in the total technology space at a certain point in time, expandability indicates the potential of the design knowledge base to further expand to unexplored technologies.

According to the formulas, the entropy is influenced by the distribution of the relevant patents (indicated by \(x_i\)) in the total technology space, and coherence and expandability are additionally influenced by the topology of the total technology space (indicated by \(\varphi_{ij}\)). Taken together, these three metrics provide a systemic assessment of the relative positions and structure of a subspace, which represents a domain’s design knowledge base, within the total technology space from different but complementary perspectives.

**STEP 5: Suggest potential technologies for the next designs**

Based on the positions and structure of a design knowledge base and the topology of the total technology space, one can further explore the unexplored technologies in the technology space for the next design opportunities. In particular, those unexplored technologies that have the highest relatedness to the established design knowledge base of a domain may present most feasible opportunities, because the high innate knowledge relatedness would enable the ease for designers in the domain to comprehend these unexplored technologies. That is, design engineers in the domain are suggested to explore engineering mechanisms, techniques and solutions from those most related unexplored technologies, and analogize, synthesize and relate them to the prior designs in their own domain for new design ideas and innovation. As a result of their exploratory efforts, the design knowledge space of the domain will be further expanded.

### 4 Case study: hybrid electric vehicles

In this section, we demonstrate the proposed methodology by applying it to analyze the design knowledge base expansion of the HEV domain and that of the leading firms in this domain. An HEV synthesizes an internal combustion engine (ICE) with an electric powertrain to combine the benefits of both ICES and electric powertrains. Due to HEVs’ advantages in fuel economy compared with traditional vehicles, the domain has attracted extensive attention, investment and design efforts from automotive companies and R&D organizations around the world. Accordingly, the design knowledge base of HEVs has
expanded substantially, especially in the past 4 decades. Varieties of HEV system configurations have been developed, and varieties of technologies have been incorporated into the design of HEVs.

### 4.1 Retrieve the patent data set for the HEV domain

We identified the US patents for technologies related to HEVs from the special patent category “903–Hybrid Electric Vehicles”, created by the USPTO, among nine art collection classes whose 3-digit IDs start with the number 9 (e.g., 901-Robotics, 977-Nanotechnology). This special category contains 2392 patents that were granted between 1914 and 2015, and assigned to 36 3-digit IPC classes. This collection of patents is used to approximate the design knowledge base of the HEV domain. Table 1 lists all the IPC classes with their respective titles, numbers of HEV-related patents and entered years, i.e., the earliest years in which an HEV-related patent was granted to each class. The patents, excluding 137 that were granted to individual inventors, were assigned to 257 assignees (firms, universities, institutes, etc.). Among the assignees, 15 firms have more than 30 HEV-related patents. 10 out of the 15 leading firms are original equipment manufacturers (OEMs), while the other

| IPC3 | Title                                | Patent count | Entered year |
|------|--------------------------------------|--------------|--------------|
| B60  | Vehicles in general                  | 2,083        | 1914         |
| F16  | Machine elements                     | 485          | 1930         |
| H02  | Electric power                        | 350          | 1930         |
| B62  | Land vehicles                        | 55           | 1934         |
| F02  | Combustion engines                   | 286          | 1965         |
| F01  | Machines or engines in general       | 80           | 1972         |
| E21  | Drilling and mining                  | 2            | 1980         |
| H01  | Electric elements                    | 58           | 1982         |
| F04  | Pumps                                | 10           | 1988         |
| B61  | Railways                             | 1            | 1993         |
| G01  | Measuring and testing                | 44           | 1994         |
| F03  | Machines or engines for liquids      | 1            | 1998         |
| F28  | Heat exchange in general             | 7            | 1998         |
| H05  | Electric techniques                  | 7            | 1999         |
| B01  | Physical or chemical processes       | 7            | 1999         |
| B64  | Aircraft                             | 5            | 1999         |
| C01  | Inorganic chemistry                  | 2            | 1999         |
| C10  | Fuels and lubricants                 | 2            | 1999         |
| G05  | Controlling and regulating           | 83           | 1999         |
| G06  | Computing                            | 160          | 1999         |
| H04  | Electric communication               | 1            | 2001         |
| B66  | Hoisting and hauling machines        | 2            | 2002         |
| F22  | Steam generation                     | 1            | 2002         |
| A01  | Agriculture                          | 2            | 2003         |
| E02  | Hydraulic and construction engineering | 5          | 2003         |
| F24  | Heating and ventilating              | 1            | 2003         |
| G08  | Signaling                            | 4            | 2003         |
| B21  | Mechanical metal working             | 1            | 2004         |
| E01  | Road, railway and bridge construction | 1          | 2006         |
| E03  | Water supply and sewerage           | 1            | 2009         |
| F17  | Storing or distributing of liquids   | 1            | 2009         |
| A47  | Furniture and appliances             | 1            | 2010         |
| E04  | Building construction                | 1            | 2012         |
| B63  | Ships                                | 1            | 2013         |
| G07  | Checking devices                     | 2            | 2013         |
| F15  | Hydraulics and pneumatics            | 1            | 2014         |
Fig. 2  Summary statistics of HEV-related patent data
are suppliers. Similarly, we approximate the HEV-related design knowledge base of each assignee using its HEV-related patent portfolio.

Figure 2a shows the cumulative patent count of the entire HEV domain (the curve on the top) and that of the leading firms in this domain over time. The curves suggest that the development of HEV-related technologies experienced a long infancy period and started to accelerate only after 1995. Figure 2b presents the cumulative number of 3-digit IPC classes holding the HEV patents in each patent portfolio over time, showing that the HEV domain and the leading firms have slowed down in the expansion into new IPC classes. In Fig. 2a, b, Toyota, Honda and General Motors (GM) have outperformed other firms since 1995. Figure 2c plots the percentage of patents assigned to each of the top ten IPC classes in each patent portfolio. The IPC class “vehicles in general” (B60) holds the most numbers of the HEV-related patents at both the domain and assignee levels, followed by “electric power” (H02), “machine elements” (F16), “combustion engines” (F02), and “computing” (G06).\(^1\)

### 4.2 Choose a technology space map for the HEV domain

To build the network map, we experimented with the four alternative knowledge relatedness measures reviewed in Sect. 2.3, namely reference-based Jaccard index, reference-based cosine index, classification-based cosine index and normalized co-classification index. Based on each measure, we calculated the relatedness percentiles of the historical new technology entries into the HEV domain when they were first covered (i.e., first HEV-related patent was classified in that technology class) between 1914 and 2015.

\(^1\) For detailed definitions of these IPC classes, please visit the website of the World Intellectual Property Organization: [http://www.wipo.int/classifications/ipc/en/](http://www.wipo.int/classifications/ipc/en/).
Figure 3 reports the cumulative probability distributions of the technology entries according to the relatedness percentiles. The dotted diagonal lines in Fig. 3 represent the probability distribution from the random scenario; new technologies are explored randomly regardless of their relatedness to the previously established design knowledge base of the HEV domain. The gap between each empirical probability curve and the dotted line indicates the extent to which the historical exploration of new HEV technologies was conditioned by the relatedness between new technologies and the prior HEV design knowledge base, based on historical patent data. Figure 3a shows the cumulative probability distributions of the 35 historical technology entries (except the starting entry B60 “vehicles in general”) of the entire HEV domain. All the empirical probability curves are steep and present statistically significant deviation from the dotted diagonal line (based on Kolmogorov–Smirnov test results reported in Table 2), suggesting the HEV design knowledge base preferentially expanded to new technologies with a higher relatedness to its prior base. Particularly, the curve corresponding to the reference-based Jaccard index is above the others, and thus indicates that it has the highest explanatory power regarding the historical expansion of the HEV design knowledge base. According to this particular curve, all of the historical technology entries but one (0.8989) has a relatedness percentile higher than 0.9.

We also ran a robustness test using weighted average relatedness to plot the curves. The reference-based Jaccard index still yields the steepest curve among alternatives. Taken together, the reference-based Jaccard index provides the highest explanatory power on the historical expansions of the HEV domain and is the relatively best choice for later analyses of this domain. Note that, this finding implies the extent of knowledge base overlapping of the technologies (as measured by the reference-based Jaccard index) strongly enables new technologies to be synthesized with the ones in the prior design knowledge base of the HEV domain. For a different domain, the most explanatory knowledge relatedness metric might be different and needs to be statistically identified again using the data from that domain.

We further confirm this choice with Kolmogorov–Smirnov tests on the deviation of the empirical distributions of the HEV domain and the leading firms from the random scenario (i.e., the dotted diagonal line). Table 2 reports the K–S statistics and corresponding p values. The positive K–S statistic values confirm the design knowledge base of the HEV domain and the firms preferentially expand to cover new technologies that are more related to the previously covered technologies and ensure the chosen relatedness metric can explain such expansions of the HEV domain and the firms.

Specifically, the reference-based Jaccard index is measured as follows:

\[
\text{Relatedness} = \frac{|C_i \cap C_j|}{|C_i \cup C_j|},
\]

where \(C_i\) and \(C_j\) denote the set of references of all patents in technology classes \(i\) and \(j\), respectively; \(|C_i \cap C_j|\) is the number of patents cited in both classes \(i\) and \(j\); and \(|C_i \cup C_j|\) represents the number of unique patents cited in either class \(i\) or class \(j\). The index value is in the range of [0, 1] and indicates the relatedness of the knowledge pieces required in the design of both technologies. If the patents in two patent classes share an identical set of references, then both technologies are designed based on a same set of knowledge pieces, and this knowledge relatedness measure is at its maximum value.

### Table 2 Results of Kolmogorov–Smirnov tests comparing the empirical distributions against the uniform distribution

| Agent            | K–S statistic | p value     |
|------------------|---------------|-------------|
| HEV domain       | 0.7905        | 0.0000E+00  |
| All assignees    | 0.8564        | 0.0000E+00  |
| Toyota           | 0.9008        | 1.7963E–13  |
| Honda            | 0.7592        | 6.5258E–09  |
| General motors   | 0.9455        | 1.3322E–15  |
| Ford             | 0.9008        | 1.8561E–09  |
| Nissan           | 0.9583        | 4.3607E–10  |
| Aisin AW         | 0.9091        | 9.3301E–09  |
| Hitachi          | 0.8632        | 4.5769E–09  |
| Denso            | 0.9569        | 2.3830E–11  |
| Suzuki           | 0.9583        | 6.0282E–06  |
| Chrysler         | 0.8408        | 8.2547E–07  |
| Daimler          | 0.9211        | 3.8229E–08  |
| Bosch            | 0.9580        | 2.6191E–07  |
| Mitsubishi       | 0.9661        | 1.0287E–10  |
| ZF Friedrichshafen| 0.9487     | 7.0934E–07  |
| Hyundai          | 0.9504        | 1.2092E–05  |

\(C_i\) and \(C_j\) denote the set of references of all patents in technology classes \(i\) and \(j\), respectively; \(|C_i \cap C_j|\) is the number of patents cited in both classes \(i\) and \(j\); and \(|C_i \cup C_j|\) represents the number of unique patents cited in either class \(i\) or class \(j\). The index value is in the range of [0, 1] and indicates the relatedness of the knowledge pieces required in the design of both technologies. If the patents in two patent classes share an identical set of references, then both technologies are designed based on a same set of knowledge pieces, and this knowledge relatedness measure is at its maximum value.
To compute the Jaccard index for each pair of the 122 patent classes, we used the richest possible historical patent data (i.e., the complete USPTO utility patent records and their citation information from 1976 to 2016) to achieve the most accurate empirical approximation of the latent technology space. The measurements conducted using the data of each of the past few decades are highly stable and consistent, in line with prior studies of patent mapping (Hinze et al. 1997; Yan and Luo 2017a). Such stability of the network map approximation of the latent technology space ensures its feasibility as a background map for longitudinal analyses of a domain within the total technology space.

### 4.3 Overlay visualization of the HEV design knowledge base in the total technology space

We now use the total technology space map to visualize the structure and expansion trajectories of the HEV design knowledge base. The original technology network built based on the 3-digit IPC classes and Jaccard index is extremely dense. Among the 7381 (= 122 × 121/2) total possible links between all pairs of technologies, only seven of them are disconnected (i.e., relatedness = 0). A visualization using such a network will not be informative. Therefore, we apply the network filtering technique introduced by Yan and Luo (2017b) to filter the original dense network. This filtered network contains only the strongest 1107 links, accounting for 15% of the total original links, but maintains 92% of the diversification paths of inventors across all links or pairs of technologies in the original full network map to explain the diversification of inventors (Yan and Luo 2017b).

In Fig. 4, the network map of the total technology space, as the background map, is overlaid with the HEV design knowledge base by highlighting the covered technologies (i.e., IPC classes in which the HEV patents are assigned) and the most likely expansion paths between them. Over the period from 1914 to 2015, the HEV design knowledge base covered 36 technologies. They are highlighted in blue and labeled with their respective 3-digit IPC class IDs, titles and entered years. The node size and color intensity, respectively, denote the total patent count and the HEV-related patent count in each class; the link width indicates the knowledge relatedness between each pair of classes. A directed purple arrow highlights the strongest (in terms of the relatedness value) link from any previously covered technologies to a newly engaged technology in the HEV design knowledge base, indicating the most likely expansion path to the new technology. In the previous step (Sect. 4.2),

Fig. 4 Total technology space map overlaid with the HEV design knowledge base and its most likely expansion paths. The map includes the IPC class ID, title and entered year of each HEV’s historical technology entry.
we have provided statistical support that the HEV design knowledge base is more likely to expand to new technologies with higher relatedness percentiles via the strongest links. The relatedness percentiles corresponding to the technology entries of the highlighted links are also reported on the map. The high values of the relatedness percentiles again suggest the domain’s strong compliance with technology relatedness in its design expansion.

The overlaid network map visually reveals that the HEV design knowledge base first emerged in 1914 from “vehicles in general” (B60, 1914), which holds the most number of HEV-related patents. In 1930, the HEV design knowledge base first expanded to cover two more technologies, “machine elements” (F16, 1930) and “electric power” (H02, 1930). Since then, two subtrees have gradually grown from these two technologies, and expanded to cover more and more new technologies, generally representing the mechanical parts and the electrical parts of HEVs, respectively. As seen in Fig. 4, the tree displays a clear gradient in node color intensities (corresponding to HEV-related patent counts) from the root through the main stems to smaller branches and then leaves. The darkest nodes appear to constitute a “backbone” in the expansion trajectories. The gradient along the tree suggests the earlier covered technologies are generally also the ones that are more frequently used in later HEV designs.

The “mechanical” subtree originating from F16 is mainly concerned with components and solutions regarding basic hybrid schemes, ICEs, transmission, lubrication and cooling systems of HEVs, such as “combustion engines” (F02, 1965), “machines or engines in general” (F01, 1972), “pumps” (F04, 1988), “heat exchange in general” (F28, 1998), “fuels and lubricants” (C10, 1999) and “machines or engines for liquids” (F03, 2009), etc. The “electrical” subtree stemming from H02 is more concerned with components and solutions regarding electric drive, sensing and controlling systems of HEVs, such as “controlling & regulating” (G05, 1999), “electric elements” (H01, 1982), “electric techniques” (H05, 1998), “measuring & testing” (G01, 1994), “computing” (G06, 1999), “electric communication” (H04, 2001), and “signaling” (G08, 2003), etc. In addition, the HEV domain also briefly branched to three technologies outside the two main trees. “Land vehicles” (B62, 1934) is highly related to the root “vehicles in general” (B60) and has accumulated 55 HEV patents. Meanwhile, “railways” (B61, 1993) and “hoisting and handling machine” (B66, 2002) only host one or two HEV patents to date.

Such overlay network visualization can also be applied to individual firms. Figure 3c and Table 2 statistically show the reference-based Jaccard index-based knowledge relatedness can explain the historical technology entries of each leading firm in the HEV domain, and the explanatory power is indeed higher for individual firms than for the domain as a whole (suggested by the steeper curves in Fig. 3c than those in Fig. 3a, b). On this basis, we choose Toyota, the leading firm in the HEV domain, to demonstrate the data-driven overlay visualization at the firm level. The HEV-related design knowledge base of Toyota is a subspace of the design knowledge base of the entire HEV domain.

Figure 5 shows the same base map as shown in Fig. 4 overlaid with only Toyota’s HEV-related design knowledge base (i.e., IPC classes in which Toyota’s HEV-related patents are assigned) and its most likely expansion trajectories. Similarly, Toyota’s HEV-related designs started first in 1974 from “vehicles in general” (B60, 1974), and then expanded to “electric power” (H02, 1995) in 1995. Over time, Toyota developed an HEV-related design knowledge base that covers 13 technologies, in two main subtrees originating from B60 and H02, and representing the mechanical parts and the electrical parts of HEVs, respectively. The tree also exhibits a declining color intensity as it continues to expand and branch out from the root. That is, the earlier-entered technologies are generally used more often in Toyota’s HEV designs over time.

In general, the data-driven visualization of the two-subtree structure in the design knowledge base of the HEV domain resonates with our basic understanding of the architecture of HEVs. The visualization also provides nuanced information on the evolutionary trajectories of the domain, i.e., the sequence of different technologies entering the design knowledge base at different times in the history and also their relative importance for the domain. Such efficient capture and systematic observations of the structure and evolutionary trajectories of a domain’s design knowledge base are made possible and easy with the data-driven visualization methodology illustrated above.

4.4 Quantitative analysis on the structure and evolution of the HEV design knowledge base

The structure and evolution of the HEV design knowledge base within the total technology space map can also be quantitatively analyzed using the entropy, expandability and coherence metrics. Specifically, the entropy increased...
rapidly during the early development stage of each design knowledge base and then entered a long steady growth stage (Fig. 6a), evidencing the increasing spread and diversity of the HEV-related technologies. The coherence of the design knowledge base also continually decreased in general (Fig. 6b), suggesting that later covered technologies are less and less related ones, but they probably only gained small weights so the change is slight at the later stage. The expandability, i.e., the potential for further expansion, continually decreased (Fig. 6c), suggesting that fewer and fewer technologies remain unexplored and they are less and less related to the design knowledge base. These trends confirm again that in the history the HEV domain preferentially expanded into more related unexplored technologies first, as suggested in Fig. 3.

Despite the generally similar trends for individual firms with that of the HEV domain as a whole, Fig. 6 also reveals the differences across firms. For instance, General Motors (GM) is the only firm whose entropy is once obviously higher than the HEV domain (Fig. 6a), suggesting that the firm has rather evenly explored diversified technologies for HEV designs. In contrast, Hyundai presents the lowest entropy (Fig. 6a) but the highest coherence (Fig. 6b) and expandability (Fig. 6c) among all firms. As a relatively new player into the HEV domain, Hyundai actually decreased the entropy and increased the coherence in recent years, despite the general entropy increase and coherence decrease of most firms and the entire HEV domain. These observations suggest Hyundai has mastered a small set of highly related technologies (i.e., low entropy and high coherence) that are also highly connected with the rest of the total technology space (i.e., high expandability), which in turn may empower its next expansion.

Denso and Daimler have the lowest coherence (Fig. 6b), because of their significant presences in such domains as “electric power” (H02) and “computing” (G06) that are relatively distant from “vehicles in general” (B60)—the dominant core of the HEV-related design knowledge bases (Fig. 2c). More exploration of the technologies in the close neighborhood of B60, such as “machine elements” (F16) and “land vehicles” (B62), would increase the coherence of

---

3 Meanwhile, the coherence of the HEV design knowledge base of the entire domain and individual firms is significantly greater than the average coherence among all 122 technologies in the total technology space (the dotted line at the bottom of Fig. 6b). This finding implies that the design knowledge base of the HEV domain and individual firms comprise a set of highly related technologies.
Fig. 6 Quantitative assessment of the network positions and structures of the design knowledge bases of the HEV domain and the leading firms over time.
the design knowledge base. Note that, Honda and Chrysler exhibit several long jumps in their HEV design knowledge base expansion histories (Fig. 3c), but their entropy and coherence are in the mid-range of those of all the firms (Fig. 6a, b), suggesting the jumps were made to technologies that were not often used in their HEV designs and thus only gained small weights in their design knowledge bases.

In the case of expandability, Honda and Toyota are the lowest (Fig. 6c). However, among the top three firms in term of HEV-related patent grants, GM consistently presents a higher entropy, expandability and coherence than Toyota and Honda. Such observations imply not only the leadership of GM to date, but also its greater innovation potentials for the future, than Toyota and Honda in the HEV domain. In brief, such analyses and findings based on the network metrics reveal the latent structural and positional characteristics of the design knowledge bases of the competing firms within the same domain.

In sum, the foregoing quantitative network analysis, together with the results in Figs. 2 and 3c, has illustrated the power of our network-based methodology by revealing the overall evolution trends of the HEV domain as well as the fundamental differences among the firms regarding the structures and evolutions of their design knowledge bases. These observations and findings are enabled by the overlay network technique that positions the design knowledge bases of the domain and the firms as subgraphs in the total technology space network, and the network-based metrics that assess the subgraphs.

4.5 Identify potential technologies for the next move of the HEV domain

Based on the overlay network maps, design engineers may further explore the unexplored technologies in the technology space for the next designs of HEV. For instance, one can start with the unexplored technologies with the highest relatedness to the established HEV design knowledge base, as a high knowledge relatedness would enable the ease for designers to exploit these technologies for new HEV design. For example, the technologies highlighted by the thick gray borders in Fig. 7 are the top 10% unexplored technologies with the highest weighted average relatedness to all the
technologies already covered in the HEV design knowledge space to the end of 2015.⁴ Then, HEV design engineers are suggested to explore engineering mechanisms, techniques and solutions from those unexplored technologies and can be leverage or relate them for HEV design. For instance, technologies in “medical & hygiene” (A61) and “conveying and packing devices” (B65) may suggest designs of HEVs for specialty applications. Many “infographics & display” (G09) technologies can be adopted for displaying the operating information of the hybrid powertrains in HEVs. Likewise, it is not difficult to imagine that technologies in “plastics working” (B29) and “machine tools” (B23) would be useful for the design and manufacturing of parts of the electrical–mechanical system of HEVs.

For Toyota, the top 10% unexplored technologies that are most related (according to weighted average relatedness) to its established HEV-related design knowledge base are highlighted in Fig. 8. Among the 11 technologies, “furniture & appliances” (A47), “aircraft” (B64), “building construction” (E04), “electric communication” (H04) and “electric techniques” (H05) have already been covered by the current design knowledge base of the entire HEV domain. Toyota may dedicate some design efforts to catch up in terms of these technologies and their applications to HEVs. Moreover, all the other technologies identified for Toyota’s future exploration have been suggested for the general HEV domain, suggesting a consistency in next moves of Toyota and the entire HEV domain. Toyota’s HEV designers may

⁴ There are alternative approaches to identify the unexplored technologies in the total technology space that are highly related to an established design space (i.e., a subspace of the total space). For instance, instead of the top 10% unexplored technologies according to their weighted average relatedness to the prior design knowledge base (approach #1), one can also identify the unexplored technologies, each of which is the most related to each of the covered technologies in the current design knowledge base (approach #2). We found approach #1 better predicted the historical technology entries at the times of their respective entries based on our HEV patent data. The hit rates of approach #1 are 49% for the HEV domain’s historical entries, 88% for all assignees’ entries together, and 71–100% for entries of individual assignees. The hit rates of approach #2 are 37% for the HEV domain, 24% for all assignees pooled together, and 0–50% for individual assignees.
also explore future design opportunities by exploring and leveraging such technologies for HEV design.

5 Discussion

5.1 Significance of the proposed methodology

In this paper, we have presented a data-driven methodology for visualizing and analyzing the design knowledge base of a technology domain using patent data. The core elements of the methodology include representing the total technology space as an empirical network map based on patent data and overlaying the network map with the design knowledge base of a domain (e.g., HEVs) or a design agent (e.g., Toyota) as a subgraph of the network. Such overlay network representations enable quantitative and visual analyses on the structure and expansion trajectories of the design knowledge base and on its knowledge relatedness to the unexplored technologies in the rest of the total technology space. We demonstrate this methodology via the case study of the HEV domain, and showcase the new types of structural, evolutionary and comparative insights that can be derived from the new methodology.

This research contributes to the growing design research and design theory literature, including the studies on infused design (Shai and Reich 2004), C–K theory (Hatchuel and Weil 2009), knowledge genomes (Reich and Shai 2012) and design-by-analogy (Fu et al. 2015), which have increasingly suggested the importance of managing knowledge across and within technology areas and exploring the knowledge space for creativity and innovation. Our work has responded to the calls from such prior design theory and methodology research by offering a data-driven knowledge management tool. The tool may potentially help researchers, engineers and managers better understand the structure and evolution of the design knowledge bases underlying their own domains or firms and also suggest unexplored technologies for new design opportunities.

The present paper may have also contributed to the prior studies on the influences of knowledge relatedness on the exploration or diversification of individual inventors, firms and cities across technological fields, by adding new empirical findings at the level of a design domain, e.g., HEVs. Specifically, the statistical analyses (Fig. 3) from STEP 2 of the proposed methodology have provided empirical evidence that the design knowledge bases of the entire HEV domain and the leading firms in this domain preferentially expanded toward new technologies that are more related to their prior design knowledge bases. We anticipate similar empirical analyses to be conducted on more and diverse design domains to test if the same pattern holds.

Note that, technologies can be defined at different levels of granularity and represented by patent classes defined at different levels (e.g., 3-digit, 4-digit, 5-digit IPC classes) for technology network mapping and analysis. In a robustness test, we found that the networks using 4-digit IPC classes yield cumulative probability distribution curves more skewed towards the relatedness percentile of 100% than those in Fig. 3 based on 3-digit IPC classes for the HEV domain. That is, the relatedness-driven expansion mechanism is even more evident for technologies of a higher granularity. In this paper, we chose 3-digit IPC classes for technology network mapping because they provide the best resolution for visual analysis, whereas there are 630 4-digit IPC classes and the network map containing all of them are cluttered and not informative. When visualizations are not required, or better visualization techniques are available, using 4-digit IPC classes for the overlay network analysis may provide finer-grained results.

In addition, although design searches in the entire technology space are relatively more global than the searchers only within the existing design knowledge base of a domain, it may still become limited as the domain matures and its design knowledge base reaches a relatively steady state (Yayavaram and Ahuja 2008). As shown in Fig. 6, with the continual expansion of the HEV design knowledge base, its expandability continually declined. It became increasingly difficult to identify and synthesize additional new technologies purely based on knowledge relatedness. In such a case, designers are more likely to identify new design opportunities by exploiting the recombination of the technologies already in the design knowledge base, which may gradually lead to search exhaustion (Fleming and Sorenson 2001; Fleming 2001; Yayavaram and Ahuja 2008). When the exploration and exploitation via knowledge relatedness reach limits in the technology space, designers may alternatively explore new technologies through more abstract types of bisociation (Berthold 2011; Thiel and Berthold 2012; Younge and Kuhn 2015), or explore design inspirations from nature (Fu et al. 2014) or the latest scientific advances (Luo 2015).

5.2 Limitations and future research

The proposed methodology should be positioned in the context of the following caveats. First of all, it is focused on the synthesis or analogy across technologies, whereas new technologies may also be inspired by nature or new scientific discoveries. Also, our methodology emphasizes the role of knowledge relatedness in incremental design knowledge base expansions, whereas radical and breakthrough innovations that arise from distant analogies, syntheses or bisociation might ignite a new domain or cause
dramatic changes in the design knowledge base of an existing domain. Therefore, it will be valuable future research avenue to develop data-driven methodologies that capture the influences of new scientific discoveries, nature-inspired designs and radical or breakthrough innovations on the emergence or expansion of design knowledge bases.

Second, as pointed out earlier in the paper, knowledge relatedness between technologies can be defined and measured in various ways, as design knowledge base expansions may be driven by different mechanisms in different contexts. In our case study, only four alternative knowledge relatedness metrics are compared for selection to illustrate our methodology framework. In future research, especially in different contexts, new relatedness measures should be explored and tested in terms of their explanatory powers. Likewise, another promising avenue for future research is to explore alternative and potentially better visualizations and structural layouts for the technology space network.

Moreover, while the case study has illustrated the power of our data-driven and network-based methodology by providing interesting observations at the domain and firm levels, the interpretations to some of the observations are limited due to the lack of the grounded firm-level data and the focus of the paper on the methodology. To better understand the interesting patterns of the HEV domain found in the case study, more in-depth data regarding the strategic choices, operations and performances of the firms need to be collected from or within the firms in the domain. Nonetheless, the observations that are enabled by our methodology have shed light on the directions for a future research project dedicated to better understand the HEV domain and the firms in this specific domain.

In addition to the deep investigation of the HEV domain, we also anticipate that the application of our methodology to more diverse technology domains and firms will yield useful insights into domain-specific practices and will enable further testing and refinement of our methodology. Particularly, the present case study shows the high explanatory power (particularly from the results in Fig. 3) of our data-driven methodology on the historical expansions of a single domain. This suggests the possibility to build a prediction model on the future expansions of generally design domains in future research. Such a prediction model needs to consider more systemic factors than knowledge relatedness and be tested with data from more diverse domains.

Acknowledgements We are grateful for the support from the Singapore Ministry of Education Tier 2 academic research Grant (#MOE2013-T2-2-167) “Theoretical Foundations of Technology Network Modeling for Innovation” and the SUTD-MIT International Design Centre. An early version of this paper was presented at the International Design Conference 2016. The feedback we received during the conference session greatly helped us refine the research and the paper.

References

Alberts D, Yang CB, Fohare-DePonio D, Koubek K, Robins S, Rodgers M, DeMarco D (2011) Introduction to patent searching. In: Current challenges in patent information retrieval. Springer, Berlin Heidelberg. pp 3–43

Allaire D, He Q, Deyst J, Willcox K (2012) An information-theoretic metric of system complexity with application to engineering system design. J Mech Des 134:100906

Alstott J, Triulzi G, Yan B, Luo J (2017a) Inventors’ explorations across technology domains. Des Sci 3:220

Alstott J, Triulzi G, Yan B, Luo J (2017b) Mapping technology space by normalizing technology relatedness networks. Scientometrics 110(1):443–479

Arthur WB (2009) The nature of technology: what it is and how it evolves. Simon and Schuster, New York

Baxter D, Gao J, Case K et al (2007) An engineering design knowledge reuse methodology using process modelling. Res Eng Des 18(1):37–48

Berthold M (2011) Bisociative knowledge discovery. Lecture notes in computer science, vol 7250. Springer, Berlin, Heidelberg, pp 1–7

Boschma R, Balland PA, Kogler DF (2014) Relatedness and technological change in cities: the rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. Ind Corp Change 24(1):223–250

Breschi S, Malerba F, Lissoni F (2003) Knowledge-relatedness in firm technological diversification. Res Policy 32(1):69–87

Chan J, Schunn C (2015) The impact of analogies on creative concept generation: lessons from an in vivo study in engineering design. Cogn Sci 39(1):126–155

D’hondt E (2009) Lexical issues of a syntactic approach to interactive patent retrieval. In: The proceedings of the 3rd BCSIRSG symposium on future directions in information access, pp 102–109

Dibiaggio L, Nasiriyar M, Nesta L (2014) Substitutability and complementarity of technological knowledge and the inventive performance of semiconductor companies. Res Policy 43(9):1582–1593

Dosi G (1982) Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. Res Policy 11(3):147–162

Ejermo O (2005) Technological diversity and Jacobs’ externality hypothesis revisited. Growth Change 36:167–195

Fleming L (2001) Recombinant uncertainty in technological search. Manag Sci 47(1):117–132

Fleming L, Sorenson O (2001) Technology as a complex adaptive system: evidence from patent data. Res Policy 30(7):1019–1039

Fleming L, Sorenson O (2004) Science as a map in technological search. Strateg Manag J 25(8–9):909–928

Fu K, Cagan J, Kotovsky K, Wood K (2013a) Discovering structure in design databases through function and surface based mapping. J Mech Design 135(3):031006

Fu K, Chan J, Cagan J, Kotovsky K, Schunn C, Wood K (2013b) The meaning of “near” and “far”: the impact of structuring design databases and the effect of distance of analogy on design output. J Mech Design 135(2):021007

Fu K, Moreno D, Yang M, Wood K (2014) Bio-inspired design: an overview investigating open questions from the broader field of design-by-analogy. J Mech Design 136(11):111102

Fu K, Murphy J, Yang M, Otto K, Jensen D, Wood K (2015) Design-by-analogy: experimental evaluation of a functional analogy search methodology for concept generation improvement. Res Eng Design 26(1):77–95

 Springer
