

Data-Driven Network Visualization for Innovation and Competitive Intelligence

Serhad Sarica¹
serhad_sarica@mymail.sutd.edu.sg

Bowen Yan¹
bowen_yan@sutd.edu.sg

Gerardo Bulato²
gerardo@innosenze.com

Pratik Jaipurkar²
pratik@innosenze.com

Jianxi Luo¹
luo@sutd.edu.sg

¹ Data-Driven Innovation Lab, Singapore University of Technology and Design

² InnoSenze Pte. Ltd.

Abstract

Technology positions of a firm may determine its competitive advantages and innovation opportunities. While a tangible understanding of the technology positions of a firm, i.e., the set of technologies the firm has mastered, can inform innovation and competitive intelligence, yet such positions are heterogeneous, intangible and difficult to analyze. Herein, we present a data-driven network visualization methodology to locate the knowledge positions of a firm as a subspace of the total technology space for innovation and competitive intelligence analytics. The total technology space is empirically constructed as a network map of all patent technology classes and can be overlaid with the knowledge positions of a firm according to its patent records. This paper demonstrates how to use the system to conduct historical, comparative and predictive analyses of the technology positions of individual and different firms. The methodology has been implemented into a cloud-based data-driven visual analytics system – InnoGPS.

1. Introduction

To compete and grow, technology firms need to continually exploit the technologies that they have mastered and also explore additional technologies for new products and services [1]-[3]. The variety of previously mastered and all unexplored technologies together constitutes the total technology space [4]-[6]. The specific set of technologies and related knowledge mastered by a firm defines its technology positions in the total technology space [7],[8]. Then, its pursuit of innovation and growth can be viewed as a process of

searching (either exploiting or exploring) and (re)combining technologies within or beyond its prior knowledge positions [3],[9],[10].

In the total technology space, the specific knowledge positions of a firm may be distant or proximate to other unexplored technology domains [11]. Two domains are proximate or distant if similar or distinct knowledge or capabilities are required to design the technologies in them. It will be relatively easy and feasible for a firm to comprehend, learn and synthesize additional technologies that are proximate to its knowledge positions than the distant ones [10],[12],[13]. Therefore, the technology positions of a firm are generally built up incrementally through a path-dependent learning process shaped by its prior positions [14],[15]. Meanwhile, exploring distant domains from a firm's prior and present capability positions may unleash the potential for radical innovation despite a higher risk and resource requirements. For growth and sustainability, firms need a balance between exploitation and exploration for both incremental innovation (economically important for the short run) and radical innovation (strategically important for the long run).

Different firms may have different intents and strategies for growth and innovation, and develop different positions. Such positions, and in particular their structural embeddedness in the total technology space, will influence the firm's future innovation potential and competitive advantages. Therefore, a tangible understanding of the technology positions of a firm within the total space of all technologies may provide innovation guidance and competitive intelligence. However, such positions of a firm are often heterogeneous, complex, intangible and difficult to understand and analyze. To address this challenge, we present a data-driven visualization methodology to support the assessment, comparison and sense-making

of the relative technology positions of firms in the total technology space for competitive intelligence.

The methodology locates the technology positions of a firm as a subspace of the total technology space, and then assess and compare the subspaces of different firms. We operationalize the conceptual “total technology space” as an empirically constructed network map of all patent technology classes defined by World Intellectual Property Office (WIPO) in line with a few recent studies on patent mapping [16]-[18]. The concept of technology positions of a firm is operationalized as the patent classes where the firm’s patents have been assigned. Specifically, we utilize the United States Patent & Trademark Office (USPTO) database to create the total technology space map and also identify the knowledge positions of a firm on the map.

In next sections, we will present the methodology after reviewing the related work, and then demonstrate the methodology via a case study.

2. Related work

A few recent studies in the information science literature have constructed network maps of patent categories [4],[16],[17] and associated these categories that are more often called as patent classes by mining the patent database. In such a network, the nodes represent technology domains and are operationalized as all patent classes defined in a patent classification system, such as the International Patent Classification (IPC), to represent technology design domains, e.g., combustion engines. The links among patent classes are weighted according to the knowledge proximity between technology domains, measured using massive patent data [11],[19]. In contrast to the patent mining and analysis works focused on small samples of patents, such network maps cover all technology classes and utilize the entire patent database to compute knowledge proximity deriving statistical significance. This mapping process helps to provide a picture of all known technologies and their relationships as complete and accurate as possible in terms of knowledge proximity (or distance). In this way such maps approximate the total technology space.

To create such maps, patent data-based measures of knowledge proximity between patent classes are required and expected to capture the intuition that the knowledge and capabilities required to design technologies in one class can also be easily used for designing technologies in the other. Some measures mine patent reference information. Jaccard index [20] can be used to calculate the count of shared references of a pair of classes normalized by the total count of all

unique references of patents in either class [21],[22]. The cosine similarity index, i.e., the cosine of the vectors of patent references made from a pair of classes to all other classes or patents, is the most popularly used metric [11],[16],[17]. Other measures mine the co-classification information, i.e., how often two classes are co-assigned to the same patents [19]. For instance, Breschi et al. [6] measured the cosine of respective patent classes’ vectors of occurrences with all other classes in patents. Dibiaggio and Nesta [21] measured the deviation of the observed co-occurrences of class pairs in patents from random expectations. Interested readers may refer to a recent review and comparison of the most popular knowledge proximity measures used in patent mapping [11].

Note that, the networks of patent technology classes are rather consistent over time [11],[24] regardless of the choices of knowledge proximity measures. For example, a recent longitudinal analysis [11] showed that the changes of all links’ weights and their relative rankings by weights over time are trivial. Such stability of the measurements and mapping may be the result of the innate but latent proximity between different physical technologies, e.g., computing and coating. That is, the proximity or distance between physical technologies has an innate physical nature; the technology space is also a latent physical existence. Therefore, the technology space approximations using data from different time periods are not supposed to vary, if sufficient data are computed and statistical significance is ensured.

In turn, the stability of the patent technology networks allows their utilization for longitudinal analysis of certain patent portfolios as subgraphs of the network map covering all patent classes. For instance, prior studies have empirically shown firms [6],[21],[25] and regions [25],[26] tend to first diversify into new technology domains near their prior domains due to the relative ease to master new but proximate technologies. Our prior analysis of 2 million inventors and 4 million patents granted from 1976 to 2010 in the USPTO (United States Patent and Trademark Office) database statistically shows that inventors are far less likely to succeed in obtaining a new patent in new technology classes that are more distant to the classes of their prior patents [10]. In another analysis of the historical patents for hybrid electrical vehicles (HEV), we found the HEV patents granted in new classes are more likely to be in those more proximate to the classes containing prior HEV patents [27].

Taken together, these prior studies on the network mapping of patent data have provided a ground for us to utilize the patent technology space map to develop a methodology and a tool to aid firms in innovation and

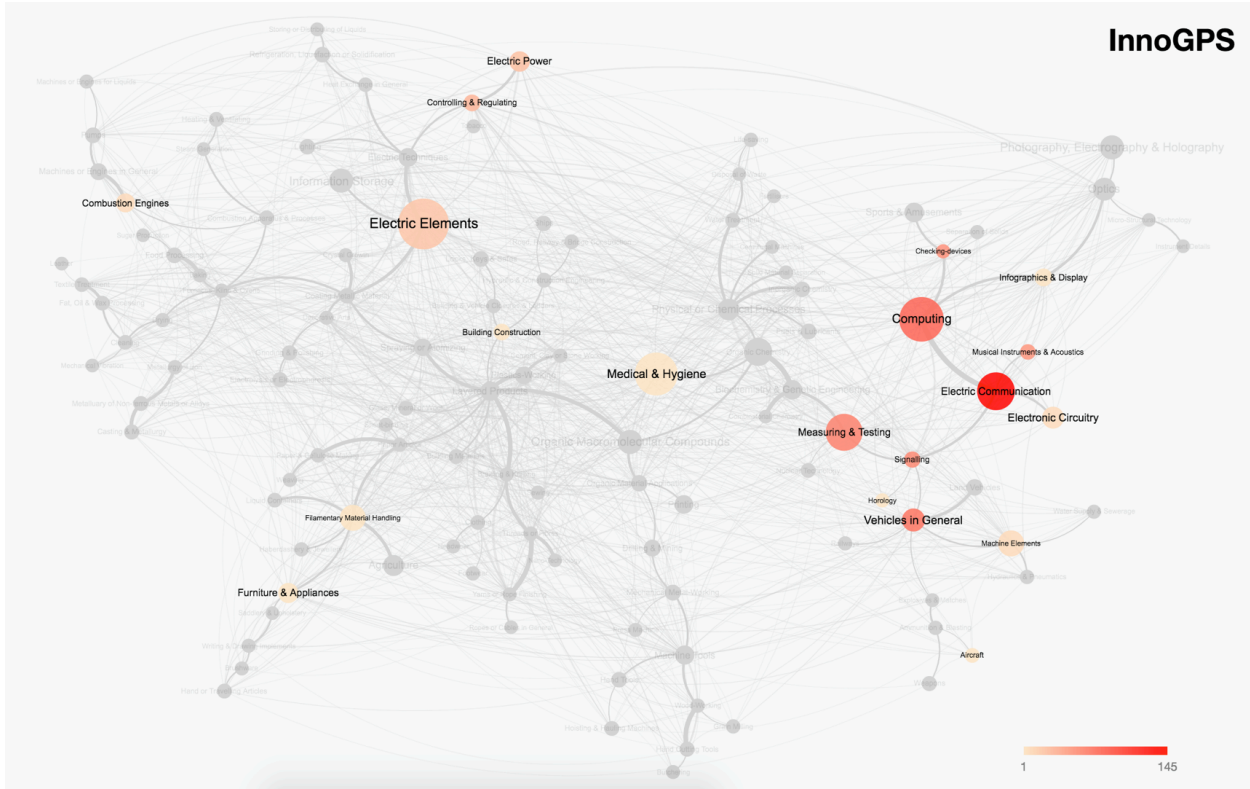


Figure 1. Technology space map of InnoGPS platform and positions of GM between 2010-2017

competitive intelligence analyses. In the following section, we will describe the methodology.

3. Method

3.1. The technology space map

The core of the methodology is the background network map that approximates the total technology space. The technology space network includes all the 122 meaningfully defined 3-digit international patent classes (e.g., combust engine, nanotechnology) to have the most comprehensive coverage of the conceptual total technology space. Although one can also use 4 to 7 digit patent classes, prior studies have suggested 3- digit classes are the most suitable and stable representations of technology domains [17] and provide the best map resolution and ease for visual analytics [18]. These patent classes as network nodes are linked and positioned according to the knowledge proximity among represented technology domains.

Herein, the knowledge proximity between each pair of technology domains is calculated as the cosine of the angle of the two vectors representing two corresponding patent classes' distributions of backward references to specific unique patents, formulated as

$$\varphi_{ij} = \frac{\sum_k C_{ik} C_{jk}}{\sqrt{\sum_k C_{ik}^2} \sqrt{\sum_k C_{jk}^2}} \quad (1)$$

where C_{ij} denotes the number of citations of all patents in patent class i to the specific patent j ; k belongs to all patents. The backward references of the patents in a technology domain approximate the knowledge base of the domain. The proximity value in [0,1] indicates the similarity of the knowledge bases of two domains.

Despite the existence of other measures of the relationships between patent categories [11],[28],[29], this metric is chosen for its high explanatory power on firms' historical diversifications across technology domains according to our statistical analysis [18]. To ensure statistical significance, more than 6 million granted patents in the USPTO database are used to calculate the knowledge proximity between each pair of the 122 technology domains. Figure 1 shows the total technology space map based on the knowledge proximity measure above. The size of a node corresponds to the total number of patents in the corresponding technology domain.

3.2. Overlay the map

For our interest in the technology positions of a firm, the total technology space map is overlaid by highlighting the subset of technology domains where the firm has been granted patents. Each patent has one or multiple patent classes. A firm may be present in one or multiple technology positions (i.e., domains) in the total technology space. For instance, the total technology space map in Figure 1 is overlaid with the technology positions of General Motors (GM) between 2010 and 2017. The red colored nodes are GM's technology positions, and the intensity of red corresponds to the portion of GM's patents in the domain over its total patents during the chosen time period. The grey colored domains do not host any patent of GM during the same time-period and represent the technology whitespace for the firm to explore for innovation, diversification and growth opportunities in future.

In turn, such an overlaid map can be used to assess the technology positions of a single firm at a time or their evolution over time and explore the next positions in future. The highlighted positions of different firms on the same background map may allow the visual comparison of different firms in terms of their technology positions for competitive intelligence.

3.3. Assessment of the technology positions of a firm

In addition to the overlay visualization, network-based metrics can be used to quantitatively evaluate the set of technology positions of a firm as a sub-network of the total technology space network. Hereafter, we consider four measures.

Entropy reveals the spread of technology positions of a firm in the total technology space or the extent of technological diversification of the firm, and takes the form of information entropy as follows

$$Entropy = \frac{1}{C} \sum_i x_i \ln(x_i) \quad (2)$$

where x_i is the portion of the firm's patents in technology domain i , and C is the maximum entropy attainable when the patents of the firm are equally distributed across all technology domains i .

Coherence concerns the proximities among the technology positions of a firm and implies the potential or the ease to recombine the mastered technologies within the firm's portfolio for new products or services in the form of incremental innovation. It is calculated as the weighted-average proximity between all pairs of technology positions of a firm.

$$Coherence = \frac{\sum_i \sum_j x_i x_j \varphi_{ij}}{\sum_i \sum_j x_i x_j} \quad (3)$$

Expandability denotes the capacity of a firm's present technology positions to be further expanded into unexplored domains (i.e., the whitespace), given their locations and connectivity in the heterogeneous total technology space. It implies the firm's potential to synthesize unexplored technologies with the ones that the firm has mastered for innovation opportunities. It can be calculated as the weighted average proximity between the firm's current technology positions and all other unexplored domains in the total space.

$$Expandability = \frac{\sum_i \sum_{j(\neq i)} x_i \varphi_{ij}}{A} \quad (4)$$

where j is the set of unexplored technology domains. A is the maximum weighted degree of nodes in technology space, and it is used to normalize the metric value into the range [0, 1].

Orientation denotes the extent of overlap of the present technology positions of the firm (i.e., the subnetwork) with the heterogeneous structure of the total technology space and indicates the level of general technology development of the firm. Its formula takes the form of soft cosine similarity as following.

$$Orientation = \frac{\sum_{i,j} \varphi_{ij} a_i b_j}{\sqrt{\sum_{i,j} \varphi_{ij} a_i a_j} \sqrt{\sum_{i,j} \varphi_{ij} b_i b_j}} \quad (5)$$

where a_i is the number of firm's patents in technology domain i and b_j is the number of all patents in domain j . Taken together, these metrics provide a systemic assessment of the technology positions of a firm from different but complementary perspectives.

3.4. The proximity of a firm's positions with unexplored domains and the positions of other firms

One can further identify the unexplored whitespace domains that are most proximate to the firm's present positions, according to the network structures around its positions. Such nearby domains host the new technologies that the firm can most easily learn, comprehend and utilize for new products and services,

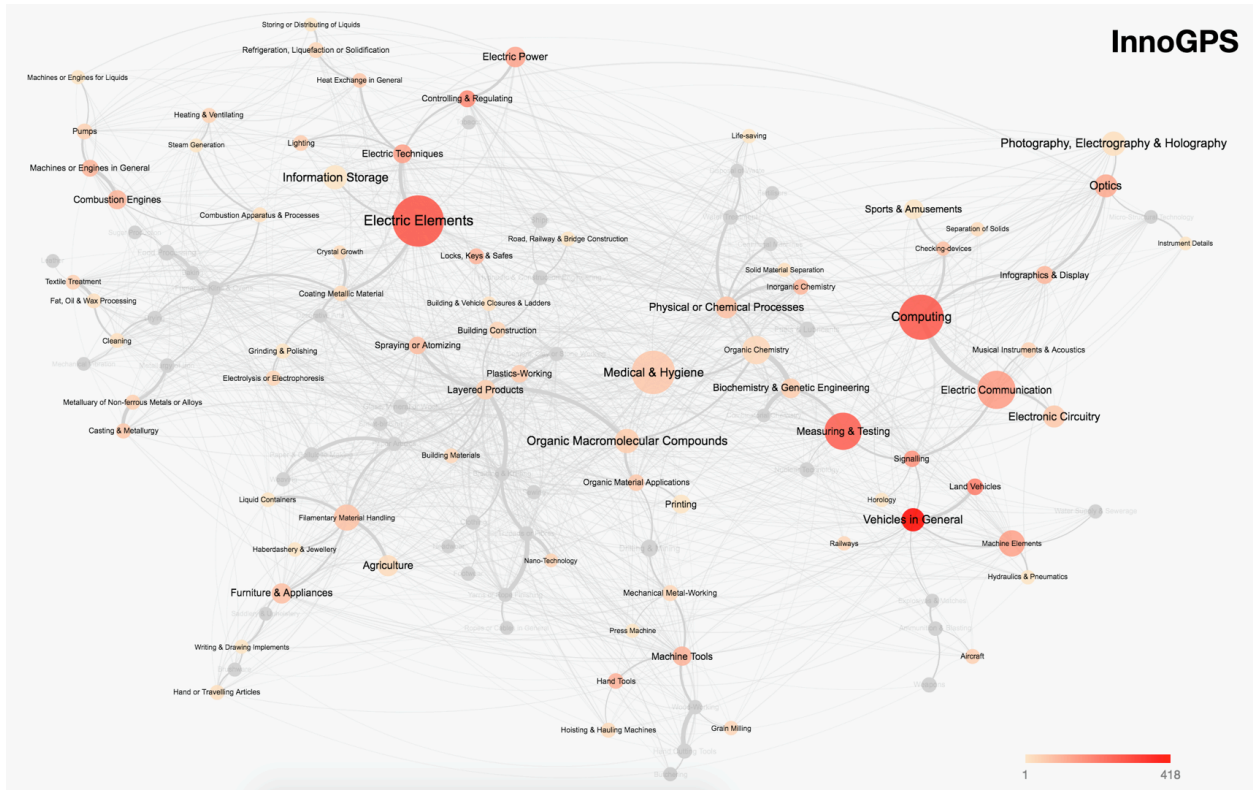


Figure 3. Positions of Toyota Motor Company between 2010-2017 in technology space map

observe such positions and monitor their changes over time.

For the GM itself, the relative proximity or distance of such positions with the white space may guide its own search for new technologies in the neighboring and even distant domains in the whitespace for next innovations. Figure 2 uses purple color and yellow circle to highlight the 5 domains in the whitespace of GM's technology positions (from 2010 to 2017) that have the highest weighted average knowledge proximity (calculated using Equation 6) to GM's current technology positions (i.e., the red nodes). These 5 most proximate domains to GM's present positions are "nano-technology", "sheet-binding", "information storage", "sports & amusements", "electric techniques". These nearby domains host the technologies that GM has yet utilized for its existing products and services, but can most easily learn, comprehend and synthesize with its previously mastered technologies for new products or services. Such ease of feasibility is enabled by high knowledge proximity between them and the technologies that GM has mastered. Therefore, these domains present near-term innovation and diversification opportunities for the firm.

In addition to single-firm analyses, the total technology space map can also serve as a benchmark for the comparison of the technology positions of different

firms. Figure 3 uses the same total technology space map as the background, but now overlays it with the technology positions of Toyota Motor Company based on the classifications of Toyota's patents during the same time-period (2010 to 2017). It is visually apparent that Toyota has more diversified technology positions including many small domains, whereas GM is only present in relatively large domains. For example, Toyota has mastered technologies in "organic chemistry", "inorganic chemistry", "physical or chemical process", "fuels and chemicals", "biochemistry & genetic engineering", "organic macromolecular compounds", etc., where GM has zero presence. Toyota is considered the global technology leader in alternative fuels and fuel cell vehicles. It is the only automotive company that has been selling proton exchange membrane fuel cell-powered sedans in the passenger car market.

"Vehicles in general" (hosting 24.3% of its patents) is still the largest technology position of Toyota (indicated by the red color intensity of the corresponding node at the bottom right area of Figure 3) despite its wide spread in the total technology space. Other strongest positions of Toyota include "computing" (9.4%), "electric elements" (8.7%), "measuring & testing" (7.5%) and "land vehicles" (5%). In contrast, GM's strongest technology position is actually "electric communication", as visually revealed

in Figure 1. In brief, the overlay map visualizations reveal the differentiation of technology positions of GM and Toyota and suggest the fundamental differences in their competitive advantages and strategies (e.g., telematics versus alternative fuels).

We apply Equation (5) to calculate the weighted proximity between the heterogeneous sets of technology positions of GM's and Toyota's. The value is 0.63, indicating only a modest similarity. These two firms that are considered head-to-head competitors have indeed a sufficient differentiation in their technological capabilities. We further calculated the weighted proximities of the technology positions of a group of automotive firms (GM, Toyota, Ford, Nissan), telecommunication firms (Cisco, Qualcomm, Ericsson, Broadcom and Huawei) and software and internet firms (Facebook, Google, Amazon and Microsoft), and used a force-directed algorithm to compute and visualize the proximity versus distance relationships among all these firms in one inter-firm network, as shown in Figure 4. The clustering structure visually reveals that GM is, in fact, more proximate with the telecommunication firms than other automotive firms. For instance, GM's technology position proximity to Huawei is 0.95, whereas its proximity to Ford Motor is only 0.4.

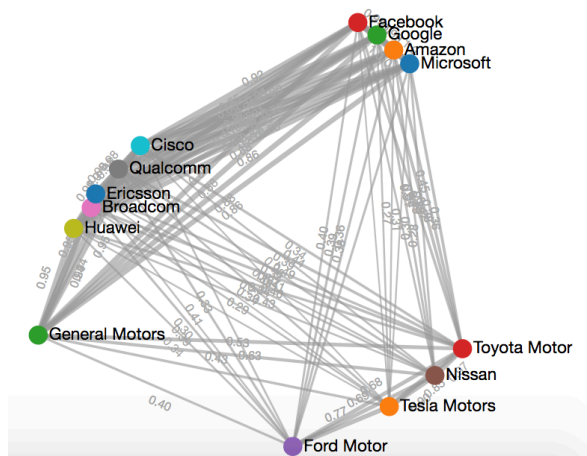


Figure 4. Knowledge proximity among the technology positions of internet, telecommunication and automotive firms. The value next to an edge is the knowledge proximity (calculated using Equation 5) between the technology positions of each pair of firms.

In addition to data-driven visual analysis, one can also utilize the network-based metrics introduced in section 3 to assess and compare the technology positions of different firms in the total technology space. Table 1 presents the entropy, coherence, expandability and orientation values of the technology positions of GM and Toyota. Higher entropy value of Toyota indicates that its technology positions are more spread out and

diversified in the total technology space. Higher coherence and expandability values of GM indicate greater prospects for next incremental innovation within the present positions and radical innovation via exploring the whitespace. The positions of both firms are similarly and modestly oriented (0.74 and 0.72) toward the general structure of the total technology space.

Table 1. Network metrics to assess and compare the technology positions of two firms (2010-2017)

Metrics	GM	Toyota
Entropy	0.43	0.66
Coherence	0.17	0.06
Expandability	0.24	0.08
Orientation	0.74	0.72

4.2. InnoGPS: Data-driven visual analytics system

The foregoing case study and multifaceted analyses showcase the sense-making power of our data-driven visualization methodology for historical, comparative and predictive analyses of firms' technology positions. This methodology, i.e., overlaying the total technology space map with a firm's technology positions based on public patent data, is aimed to enhance knowledge management, technology road mapping, and competitive intelligence analysis toward a more data-driven and visual-informed fashion.

To support relevant research and practice, we have implemented the data-driven visualization methodology in a cloud-based system called InnoGPS (www.innogps.com), which stands for "Innovation Global Positioning System", as the core function of the system is to position firms, individuals or regions in the total technology space according to their innovation records, and make use the information of such positions, relative to the white space, in the total technology space, for innovation and competitive intelligence. The functions and design of InnoGPS are largely analogous to the GPS that we use for positioning, neighborhood exploration, and direction finding in the physical space. Figure 5 is a screenshot of InnoGPS.

In addition to the function to visually position a firm in the map and quantitatively analyze heterogenous network positions, the system also reports detailed information of the firm's inventors and patents in each of the technology positions of the firm, which is valuable for competitive intelligence analytics. The inter-firm knowledge proximity analytics as illustrated in Figure 4 are also automated in InnoGPS. For innovation intelligence, InnoGPS requires only one click to highlight the white space domains according to their proximities to the firm's positions and recommend

network routes from the firm's current positions to any chosen white space domain for the planning of incremental technology capability building into the future.

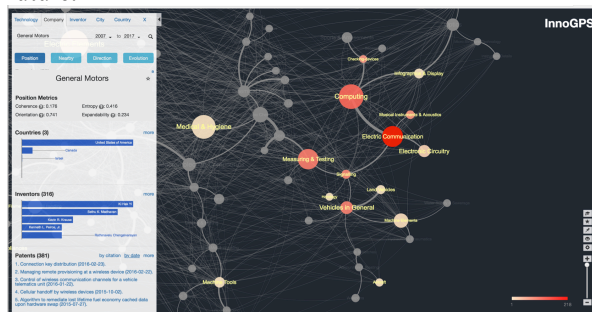


Figure 5. InnoGPS platform

5. Summary

This paper presents a new data-driven visualization methodology and a system to assess and compare the technology positions of firms in the total technology space for competitive intelligence. The methodology is based on the synthesis of innovation theories, network analysis and visualization, information sciences, and patent data. The GM-Toyota comparative case demonstrates its power for visual sense making. Its implementation in the cloud-based InnoGPS system is aimed to provide firms and managers with rapid, data-driven, scientifically grounded, and visually-informed innovation opportunities and competitive intelligence.

Moving forward, the utility and effectiveness of our data-driven methodology and the InnoGPS are still conditioned on a few maps and system design factors and need to be further tested and exploited. For instance, 4- to 7-digit classes as domains may provide more nuanced information, whereas 3-digit classes may provide analytics at a more macro level. Furthermore, micro-level analysis of technology classes may reveal indications of varying relations in technology space. Alternative knowledge proximity measures and visualization techniques may lead to different network structures and map layouts, affecting the human understanding of the visual results. Evolutionary analysis of network metrics may provide indications on the innovation strategies and development stages of firms. Future research is required to explore and experiment alternative map construction strategies and network analysis methods.

The general framework constituted by the data-driven methodology can be modified to various kinds of maps for different mediums. One such medium can be science domain, where a map of scientific fields can be constructed using enormous scientific publications data available to position research institutions on the space of science.

In conclusion, we hope this research may invite more uses of the data-driven visualization methodologies in both research and practices for innovation and competitive intelligence.

6. References

- [1] Schumpeter, J. A. (1934). *The theory of economic development: an inquiry into profits, capital, credit, interest, and the business cycle*. New Brunswick, New Jersey: Transaction Books.
- [2] Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research Policy* 11(3): 147-162.
- [3] March, J. G. (1991) Exploration and exploitation in organizational learning. *Organization Science* 2 (1): 71–87
- [4] Alstott, J., Triulzi, G., Yan, B., and Luo, J. (2017a) Mapping technology space by normalizing patent networks. *Scientometrics* 110(1): 443-479.
- [5] Teece, D. J., Rumelt, R., Dosi, G., Winter, S. G. (1994) Understanding corporate coherence: theory and evidence. *J Econ Behav and Organ* 23(1): 1–30.
- [6] Breschi, S., Malerba, F., Lissoni, F. (2003) Knowledge-relatedness in firm technological diversification. *Research Policy* 32(1): 69–87.
- [7] Yayavaram, S. and Ahuja, G. (2008) Decomposability in knowledge structures and its impact on the usefulness of inventions and knowledge-based malleability. *Admin Science Quarterly* 53(2): 333-362.
- [8] Dibaggio, L., Nasiriyar, M., Nesta, L. (2014) Substitutability and complementarity of technological knowledge and the inventive performance of semiconductor companies. *Research Policy* 43(9): 1582-1593.
- [9] Fleming, L. (2001) Recombinant uncertainty in technological search. *Manage Science* 47(1): 117-132.
- [10] Alstott, J., Triulzi, G., Yan, B., Luo, J (2017b) Inventors' explorations across technology domains. *Design Science* 3: e20.
- [11] Yan, B., and Luo, J. (2017a) Measuring technological distance for patent mapping. *Journal of the Association for Information Science and Technology* 68(2): 423-437.
- [12] Ward, T. B. (1998) Analogical distance and purpose in creative thought: Mental leaps versus mental hops. *Advances in analogy research: Integration of theory and data from the cognitive, computational, and neural sciences*.
- [13] Srinivasan V., Song, B., Luo, J., Subburaj, K., Rajesh, M. E., Blessing, L., Wood, K. (2018) Does analogical distance affect performance of ideation. *Journal of Mechanical Design*, In Press.
- [14] Winter, S. G. (2000) The satisficing principle in capability learning. *Strategic Management Journal*, 981-996.
- [15] Nelson, R. R., and Winter, S. G. (2009) *An evolutionary theory of economic change*. Harvard University Press.
- [16] Kay, L., Newman, N., Youtie, J., Porter, A. L., and Rafols, I. (2014) Patent overlay mapping: Visualizing

- technological distance. *Journal of the Association for Information Science and Technology* 65(12): 2432-2443.
- [17] Leydesdorff, L., Kushnir, D., and Rafols, I. (2014) Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC). *Scientometrics* 98(3): 1583-1599.
- [18] Yan, B., and Luo, J. (2017b) Filtering patent maps for visualization of diversification paths of inventors and organizations. *Journal of the Association for Information Science and Technology* 8(6):1551-1563.
- [19] Engelsman, E. C., and van Raan, A. F., (1994), "A patent-based cartography of technology," *Research Policy*, 23(1), pp. 1-26.
- [20] Jaccard, P., (1901), "Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines," *Bull. Soc. Vaud. Sci. Nat.*, 37, pp. 241-272.
- [21] Von Wartburg, I., Teichert, T., and Rost, K. (2005), "Inventive progress measured by multi-stage patent citation analysis," *Research Policy*, 34(10), pp. 1591-1607.
- [22] Leydesdorff, L., and Vaughan, L., (2006), "Co-occurrence matrices and their applications in information science: extending ACA to the web environment," *Journal of the American*
- [23] Dibiaggio, L., and Nesta, L., (2005), "Patents statistics, knowledge specialisation and the organisation of competencies," *Revue d'économie industrielle*, 110(1), pp. 103-126.
- [24] Hinze, S., Reiss, T., and Schmoch, U., (1997), Statistical analysis on the distance between fields of technology, Fraunhofer-Inst. Systems and Innovation Research.
- [25] Boschma, R., Balland, P.-A., and Kogler, D. F. (2014), Relatedness and technological change in cities: the rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010, *Industrial and Corporate Change*, 24(1), 223-250.
- [26] Rigby, D. L., (2015) Technological relatedness and knowledge space: entry and exit of US cities from patent classes, *Regional Studies*, 49(11), pp. 1922-1937.
- [27] Song, B., Triulzi, G., Alstott, J., Yan, B., and Luo, J., (2016) Overlay patent network to analyze the design space of a technology domain: The case of hybrid electrical vehicles. *Proc. DS 84: Proceedings of the DESIGN 2016 14th International Design Conference*.
- [28] Joo, S., Kim, Y. (2010) Measuring relatedness between technological fields. *Scientometrics* 83: 435-454.
- [29] Altuntas, S., Dereli, T., Kusiak, A. (2015) Analysis of patent documents with weighted association rules. *Technological Forecasting and Social Change* 92: 249-262.