http://dx.doi.org/10.1080/09537325.2016.1277202

Identifying R&D partners through Subject-Action-Object semantic analysis in a problem & solution pattern

Xuefeng Wang[a], Zhinan Wang[a], Ying Huang[a], Yuqin Liu[b], Jiao Zhang[a], Xiaofan Heng[a] and Donghua Zhu[a]

[a]School of Management and Economics, Beijing Institute of Technology, Beijing, People's Republic of China; [b]Academy of Printing and Packaging Industrial Technology, Beijing Institute of Graphic Communication, Beijing, People's Republic of China

ABSTRACT

Today's companies still rely heavily on expert knowledge rather than quantitative data with a systematic approach to effectively identify and choose Research and Development (R&D) partners. It is advantageous to identify and select potential R&D partners using a Problem & Solution (P&S) pattern. This paper presents a novel process for identifying R&D partners on the basis of solution similarities that assist technology managers in understanding the relationships between research targets. First, we choose a thematic dataset that contains problems and quantitative data with relative topic terms. Then, we extract Subject-Action-Object semantic structures in a P&S pattern from the dataset, and identify various solutions to a technical problem, with each as a subject. In addition, we provide correlation mapping to visualise the text characters and identify R&D partners. Finally, we validate the proposed method through a case study of the dye-sensitized solar cells sector.

ARTICLE HISTORY
Received 9 March 2016
Revised 5 December 2016
Accepted 22 December 2016

KEYWORDS
Partner identification; Subject-Action-Object semantic analysis; term clumping; correlation mapping; problem & solution pattern; dye-sensitized solar cells (DSSCs)

Introduction

Since the 1980s, this competitive environment has been increasingly unstable, and the life cycle of products and technology continues to diminish. Firms are forced to reconsider their Research and Development (R&D) configuration to manage their technology base more effectively (Nijssen, Van Reekum, and Hulshoff 2001). Hence, cooperative R&D plays an important role in technological innovation, especially in an open innovation model. Cooperative R&D is a vehicle for firms to learn skills and capabilities from their partners. Moreover, cooperative R&D agreements can be used by firms to set rules in settings with high costs and high risks (Ahuja 2000; Hagedoorn 2002; López 2008). Firms have gradually modified their innovation network by including more external partners operating outside their core areas (Mowery, Oxley, and Silverman 1998). Alliances are playing an increasing role in open innovation, thus supporting the idea that firms are intensely searching for weak ties between their innovation process and external factors in a typical open innovation approach (Chesbrough 2006; Laursen and Salter 2006). As mentioned above, open innovation is a highly debated issue, where we should recognise the importance of enterprises' independence in risk sharing and sharing intellectual capabilities. Cooperative R&D should emphasise the progress of scientific and technological knowledge. Different types and forms of knowledge contribute to technology development such as formal codified knowledge, tacit knowledge, informal knowledge, and cultural

CONTACT Xuefeng Wang wxf5122@bit.edu.cn
This article was originally published with errors. This version has been corrected. Please see Corrigendum (http://dx.doi.org/10.1080/09537325.2017.1283871).

knowledge (Fleck 1997). Improved partner identification and selection reduce randomness and uncertainty in technology development.

Many factors influence the performance of R&D partner selection; for instance, technological similarity and complementarity, developmental strategy, cultures of enterprises, top manager attitude (Chen et al. 2010), and R&D cooperation consciousness (Wang 2012). The effect of R&D collaboration may vary depending on partner types (Kang and Kang 2010). Enterprises that want to enhance their existing technical capacity or share costs and risks should select potential partner targets that have similar techniques. Enterprises that want to learn new skills and improve their technology weaknesses or fill in blank areas should look for potential targets with complementary solutions to their problems.

Over the past few decades, social science relies more and more on the combination of qualitative and quantitative approaches (Teddlie and Tashakkori 2009). Although some research using quantitative methods has been done on partner selection (Solesvik and Encheva 2010), there has been little focus on using systematic semantic analysis methods to identify partners. Many scholars have provided insight. Brouthers, Brouthers, and Wilkinson (1995) answered the question, 'When should strategic alliances be chosen?' Hagedoorn (2002) gave an overview of inter-firm collaboration or strategic partnering. Wu, Shih, and Chan (2009) proposed the integrated approach of an Analytic Network Process (ANP) for partner selection criteria in strategic alliances, and Baum, Cowan, and Jonard (2010) argued that knowledge complementarity might be the true causal force behind alliance formation. Park et al. (2015) provided a new systematic methodology to explore potential R&D collaboration partners using patent information. However, these methods can only be used in the case of known partner candidates, and the relationships were mined on the basis of frequencies of bibliographic items. These frequencies were calculated using bibliometric methods such as International Patent Classification (IPC) analysis (Angue, Ayerbe, and Mitkova 2014) and co-citation analysis (Lai and Wu 2005). However, bibliometric methods can only determine similar literature articles on the basis of fields such as keywords, IPC codes, and citations, which do not reflect technology content. Subject-Action-Object (SAO) can explicitly describe the functional relationships among components used in technological text.

Being aware of the numerous factors that can influence partner selection, we propose a systematic process for identifying partners. This paper attempts to answer the following two questions: (1) How can an organisation more effectively measure the similarity of each research target's technology on the basis of patent information? (2) How can an organisation identify potential partners based on the SAO semantic structures of different research targets? We address the process of identifying potential partners by combining term clumping (Zhang et al. 2014a), which enables users to quickly extract meaningful topic terms from large amounts of text and SAO semantic analysis in a Problem & Solution (P&S) pattern (Kim et al. 2009). This paper draws on Science, Technology, & Innovation (ST&I) data. Using traditional text mining techniques and term clumping on the ST&I data, we draw out topic terms that are highly relevant to the technical problem. However, these terms cannot help us find a solution to a specific technical problem. We extract SAO structures nearby the topic terms. For a specified technical problem (object), all SAO structures extracted have the same AO (action & object), so we can construct correlation mapping of research targets on the basis of a subject as the technical solution. From the mapping network, we can identify potential partners.

The paper is structured as follows. The Literature Review section gives a brief review of SAO semantic analysis and R&D partner selection. The Methodology section introduces the process of potential partner identification. Then, in the Case Study: DSSCs Sector, we provide empirical results of this method using a case study. Finally, the Conclusion section gives a summary and indicates promising research opportunities to pursue in the future.

Literature review

R&D partner selection

Several studies show that firms use external partnerships to face technological changes and the increasing complexity of knowledge processes (Hill and Rothaermel 2003; Keil et al. 2008; Noseleit and de Faria 2013). Firms rely on science-based partnerships to experiment with new technologies, as well as to refine existing technologies (Cohen, Nelson, and Walsh 2002). R&D partners can get access to not only tacit scientific knowledge, but also to codified knowledge, allowing them to quickly build on the latest research findings (Fabrizio 2009; Yoon and Song 2014). As choosing technology partners is a multi-criteria decision-making task, a defined selection criteria and methodology should be used. A partner's culture, past experience, size, and structure are as important as task-related factors, such as financial assets, managerial experience, and access to markets. Yoon and Song (2014) summarised methods for partner selection into three categories: mathematical programming approaches (Solesvik and Encheva 2010), rating/linear weighting approaches (Wang and Chen 2007), and artificial intelligence techniques (Fischer, Jähn, and Teich 2004). Solesvik and Westhead (2010) wrote that the following quantitative methods could be used to select partners: the ANP (Chen, Lee, and Wu 2008; Sarkis, Talluri, and Gunasekaran 2007; Wu, Shih, and Chan 2009), the analytic hierarchical process (Mikhailov 2002), optimisation modelling (Cao and Wang 2007), and the goal programming technique (Hajidimitriou and Georgiou 2002). Different researchers divided partner selection into different phases. Samadhi and Hoang (1998) summarised the process of partner selection by dividing it into three phases: scanning potential partners, matching partners for compatibility, and logistic considerations. Talluri, Baker, and Sarkis (1999) proposed a two-phase quantitative framework to aid the decision-making process of effectively selecting an efficient and a compatible set of partners. In addition, there are also some other studies that delve into the depths of the impact of partners' technological relatedness on inter-organisational links (Cloodt, Hagedoorn, and Van Kranenburg 2006; Lane and Lubatkin 1998). Technological relatedness refers to the comparison of partners' basic knowledge, which, if sufficiently close, illustrates a form of affinity to their technological knowledge (Angue, Ayerbe, and Mitkova 2014). Many organisations are selling their technological knowledge (Bianchi et al. 2011) and participating in the rise of a genuine market for technologies (Angue, Ayerbe, and Mitkova 2014; Gambardella and McGahan 2010). Alliances with partners with similar characteristics might have a positive impact on innovative performance (Mowery, Oxley, and Silverman 1996). The approach in this paper provides information for the initial phase-scanning potential partners using technological knowledge.

SAO semantic analysis

Text mining helps a great deal by extracting terms and multi-terms, their frequencies, and other dimensions, but the method lacks in semantic relationships. A researcher needs to determine the relationship between concepts in order to discover any potential concepts (Vicente-Gomila 2014). SAO emphasises key concepts. The structure explicitly describes the relationship among different phrases containing different semantic information. First, it can represent functions of technology; for example, 'battery energises bulb' represents the function of a battery. Second, it can describe a relationship between components (Cascini, Fantechi, and Spinicci 2004). Subjects and objects might refer to components of a system, whereas actions might refer to functions performed by and on components, for example, 'porous membrane (S) contains (A) metal oxide (O)'. Third, it can state partitive relationships among products or technologies. Fourth, it can be organised in a P&S format (Moehrle et al. 2005). We estimate the relationship with the P&S pattern in which Action and Object (AO) indicate the problem and subject (S) represents the solution (see Figure 1).

In 2004, Verbitsky (2004) proposed the concept of *semantic TRIZ*, which was a novel approach to the innovation process; it applied semantic indexing technology to the traditional problem-solving TRIZ theory. Kim et al. (2009) defined the basic tasks of extracting problem and solution key phrases that constitute a technology and technological trends and proposed a Technological Trend Discovery system that can automatically capture technological mainstream terms from thousands of related documents. Zhang et al. (2014b) emphasised semantic TRIZ approach as a useful tool to process 'Term Clumping' results to retrieve P&S patterns and applied them to Technology Roadmapping. With the development of the SAO semantic analysis, it has been used in many sides, such as

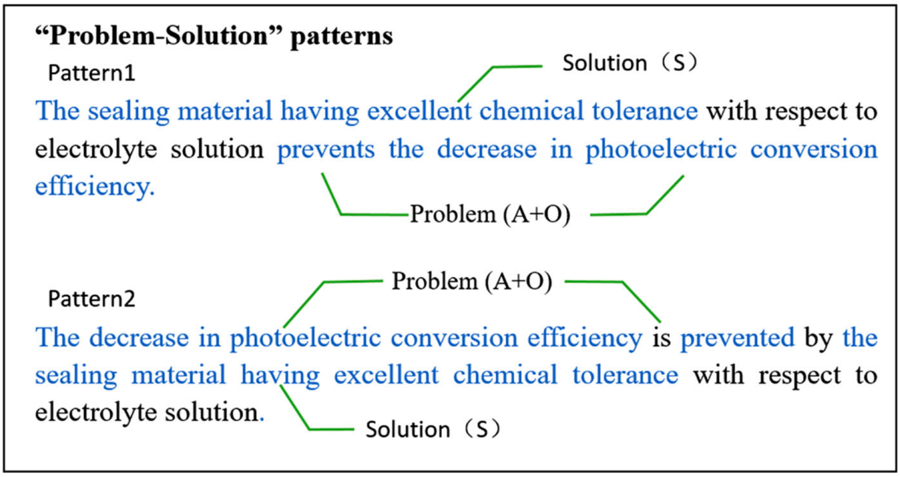

Figure 1. SAO structures in a P&S pattern.

monitoring technology (Gerken and Moehrle 2012), identifying technology development trends (Wang et al. 2015), determining the direction of technological change (Guo et al. 2016), generating patent maps, constructing a technology tree for technology planning (Choi et al. 2012b), identifying patent infringement (Bergmann et al. 2008), creating a function-based technology database (Choi et al. 2012a), and identifying technological competition trends for R&D planning (Yoon, Park, and Kim 2013).

Methodology

The SAO structure can reflect R&D substance effectively and then reveal the relationship between technical problems and solutions. We transform the research targets' correlation based on solution terms into correlation based on the *S* structure. This transformation can effectively reflect the homogeneity or heterogeneity among targets based on their research contents. From there, we can efficiently identify an R&D partner. For this purpose, we propose a process of potential R&D partner identification as shown in Figure 2.

Thematic dataset construction based on term clumping

Patent data are valuable in exploring the opportunities of promising R&D development because patents are the output of R&D and are vast public resources that contain technical and market value. To gain high-quality patent data, we need to construct an effective retrieval strategy composed of logical query formation, including a title with rich content, an abstract, keywords or subject terms, as well as the publication/application time, IPC, and so on. In addition, through literature reviews and tech mining (Porter and Cunningham 2004), we can refine technical problems in a specific technology domain.

We aim to handle challenging analyses of millions of phrases and terms derived from ST&I datasets using Natural Language Processing (NLP) with term clumping. Term clumping is the method used to clean and cluster rich sets of topical phrases and terms derived with NLP techniques from a collection of technical documents specific to one technical domain. This method minimises noise and maximises prominent topics, which enables users to extract meaning from large amounts of text more quickly. We use the professional text mining software VantagePoint (www.theVantagePoint.com) to extract terms and phrases that provide more effective information for

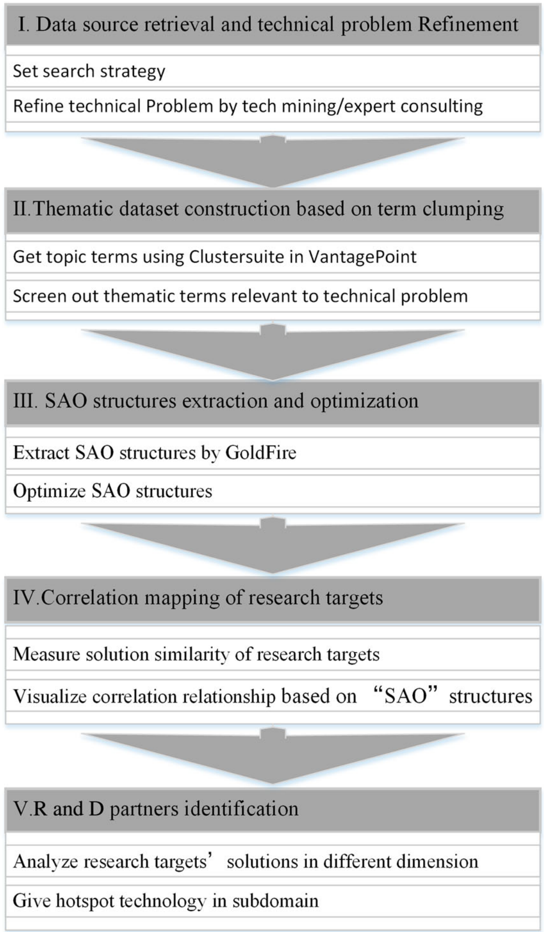

Figure 2. R&D partners' identification process in a P&S pattern.

follow-up analysis. Term clumping combines a number of textual analysis techniques, such as stop word lists and synonym list construction, fuzzy set matching, Term Frequency-Inverse Document Frequency (TFIDF), and Principal-Components Analysis (PCA). ClusterSuite is an application written in the VBA programming language with an HTML user interface that runs as a script within VantagePoint. It contains three phases: Phase I is currently the most developed phase. It executes five thesauri and one list-cleaning macro. Phase II runs one of two term clumping macros. Phase III is designed to eliminate extreme list components on the basis of parameters input by the user through the HTML interface. The end goal of ClusterSuite is to be an efficient, user-friendly application to assist in clustering terms and phrases that were previously unrelated. In the end, we have a thesaurus and a constructed dataset with more related thematic terms. Depending on the specific technical problem, we choose terms with high correlation from the thesaurus and construct the thematic dataset.

Correlation mapping of research targets in a P&S pattern

We utilise the GoldFire software and manual operation as the main tools for extraction. The SAO structure can be extracted from the title, the abstract, and the full text of technical literature. We retrieve data from these works on the basis of the thematic terms in the term clumping step. The nearby verbs are the actions that connect the problem and the solution. We pay more attention to the system components but not to the action. Considering the incomplete nature of NLP, we filter the unnecessary verb phrases and SAO structure with impurities on the basis of manual processing. In a P&S pattern, the subject that can solve the problem serves as the solution of each target.

We first divide the SAO structure positioned as a solution into three dimensions, such as materials, technologies, and components. By analysing the co-occurrence matrix between research targets (T) and the solution (S), we utilise the TFIDF method to indicate the text feature of each solution. The following vector cosine is chosen to measure the correlation degree of research targets:

$$\mathrm{Sim}(T_i, T_j) = \cos(\theta) = \frac{\sum_{k=1}^{n} w_{ik} w_{jk}}{\sqrt{\sum_{k=1}^{n} w_{ik}^2} \times \sqrt{\sum_{k=1}^{n} w_{jk}^2}}. \tag{1}$$

Here, $T_i = (w_{i1}, w_{i2} \cdots w_{in})$ $T_j = (w_{j1}, w_{j2} \cdots w_{jn})$, w_{ik} $(k = 1 \ldots n)$ is the TFIDF value of the solution set S={$S_1 \ldots S_n$} in the research target T_i. Then, we map the correlation relationships using network visualisation. Considering the display effect, we then number the research targets, which may be other companies, universities, or research institutes.

R&D partners' identification

R&D refers to future-oriented, long-term activities in science or technology. Research targets have different R&D motivations and strategies in different time periods. We use the map we construct to position each target's solution and cooperation, analyse the similarities in their research content, and tap the potential cooperation opportunities to identify potential partners among them. Potential R&D collaboration partners are visualised as a patent assignee-level map on the basis of the technological similarity between solutions, using network analysis. This stage can be divided into two steps. Based on a specific technical domain, we first focus on subdomains and then look at each subdomain. Targets can be aware of their potential partners' solutions and hotspot technology within a subdomain. If organisations want to improve their professional strength in an original subdomain, they can identify partners in the same subdomain. Organisations can also see targets that can enhance comprehensive strength in another subdomain. In addition, they can see the hotspot technology of a whole domain that may be relevant to a specific technical problem. The map results in an R&D cooperation programme for the chosen targets.

Case study: DSSCs sector

Compared to other fossil fuels, DSSCs cause less pollution, involve low manufacturing cost, and provide high-energy conversion, which cause a wide attention all over the world. Our team has focused on DSSCs research for several years. Previously, we learned that the DSSCs domain has three main technical problems: how to enhance the efficiency of the photoelectric conversion; how to reduce the cost of battery production; and how to reduce the pollution of the chemistry (Wang et al. 2014, 2015). In this paper, DSSCs are chosen as the case, the photoelectric conversion efficiency is the technical problem, and the Derwent Innovation Index database is the source data. On the basis of the above identification process, we explore the R&D partner identification and potential opportunities of DSSCs domain organisations.

Data retrieval and technical problem refinement

It is widely accepted that the quality of datasets strongly influences the analysis results; an ideal dataset is dependent on an accurate retrieval strategy (Huang et al. 2015). As mentioned above, we have focused on DSSCs for several years, we use the refined search strategy after amendments and improvements (Huang et al. 2016). The time period investigated in this paper is 1991–2014. We chose 1991 as the initial year because the earliest source article about DSSCs was published in

1991 in *Nature* by O'Regan and Gratzel (1991). Finally, we accessed 7003 pieces of patent data 16 May 2015. Note that patents from the year 2014 might be absent because of the lag of the patent disclosure cycle.

Thematic dataset construction based on term clumping

As detailed previously, term clumping was used to obtain keywords associated with the technical problem, 'How to enhance the efficiency of the photoelectric conversion.' Via the tool *ClusterSuite* in VantagePoint, Table 1 presents each procedure and results; in the end, we had 2734 keywords.

We read the above 2734 keywords manually and consulted experts' opinions in relevant fields, finally screening out 164 highly relevant topic terms that applied to the technical problem of conversion efficiency, which appeared in 3363 patents. These keywords can be divided into two categories: (1) those that directly enhance conversion efficiency, such as 'high photoelectric conversion efficiency', 'high-energy conversion efficiency', 'excellent photoelectric conversion characteristics', etc.; and (2) those that indirectly enhance conversion efficiency and include enhanced electronic conductivity (such as 'excellent electro-conductivity', 'high electrical conductivity', 'excellent ionic conductivity', etc.), enhanced catalytic activity (such as 'high catalytic activity', 'excellent photocatalytic activity', 'high photocatalytic efficiency', etc.), and reduced energy consumption (such as 'low energy consumption', 'reduced energy consumption', 'energy-saving', etc.). To an extent, the number of patents can reflect the technological strength of an organisation; these organisations are the main partner targets in R&D cooperation. For this purpose, the paper chose the top 30 organisations as the research targets, which have a total of 1423 patents. Table 2 shows the top 30 organisations.

Correlation analysis based on SAO structures extracted

Often the titles and abstracts of articles can reflect the whole idea well. In consideration of the extensive abstract content, we eventually extracted the SAO structure from patent abstracts with the help of the GoldFire software. In the end, we had 559 SAO structures related to 'enhance conversion efficiency' as shown in Table 3. Based on the literature review for DSSCs and the experts' aid, we distinguished four subdomains of DSSCs in the dimension of the component defined as *photo anode*, *sensitizer*, *electrolyte*, and *counter electrode* (Zhang et al. 2014a). We divided the 559 SAO structures by subdomain, which had 304 structures about *photo anode*, 151 about *sensitizer*, 85 about *electrolyte*, and 19 about *counter electrode*. Obviously, *photo anode*, *sensitizer*, and *electrolyte* are the main research topics.

Of the 559 SAO structures, each organisation has a solution in the *photo anode* subdomain. In the *sensitizer* subdomain, the following organisations do not have solutions: C6, C7, C14, C17, C18, C22,

Table 1. Term clumping results.

Stage	Field selection	Title and abstract (NLP phrases + keywords)
Applying thesauri for common term removal	Phrases with which we begin	91,769
	NumPunctToSpace. the	83,363
	XMLEncoding. the	83,363
	Chemical_Compounds. the	81,687
	Common_and_Basic. the	78,758
	Scientific_and_Academic. the	78,052
	Dssc-thesaurus.the	61,886
Pruning	High Extremes Removed	61,885
	General-fuzzywordmatch-1 exact.fuz	59,626
	Low Extremes Removed	14,792
Screening	Combine Author Networks	2734

Table 2. Assignee codes of organisations.

No	Assignee code	Organisation name	No	Assignee code	Organisation name
C1	KONS-C	KONICA CORP	C16	YAWH-C	NIPPON STEEL CHEM CO LTD
C2	FUJF-C	FUJI FILM CORP	C17	KYOC-C	KYOCERA CORP
C3	FUJD-C	FUJIKURA LTD	C18	ETRI-C	ELECTRONICS & TELECOM RES INST
C4	SHAF-C	SHARP KK	C19	TOXW-C	TOYO INK MFG CO LTD
C5	SMSU-C	SAMSUNG SDI CO LTD	C20	SEKI-C	SEKISUI CHEM IND CO LTD
C6	SONY-C	SONY CORP	C21	DENK-C	TDK CORP
C7	NIPQ-C	DAINIPPON PRINTING CO LTD	C22	KOAD-C	KOREA INST SCI & TECHNOLOGY
C8	NIIT-C	NAT INST ADVANCED IND SCI & TECHNOLOGY	C23	NIPK-C	NIPPON KAYAKU KK
C9	DONG-N	DONG JIN SEMICHEM CO LTD	C24	TOKE-C	TOSHIBA KK
C10	TOYW-C	TOYOTA CHUO KENKYUSHO KK	C25	CHSC-N	CHINESE ACAD SCI SHANGHAI CERAMICS INST
C11	GLDS-C	LG CHEM LTD	C26	TOPP-C	TOPPAN PRINTING CO LTD
C12	AISE-C	AISIN SEIKI KK	C27	IRIC-N	IRICO GROUP ELECTRONICS CO LTD
C13	MITY-C	MITSUBISHI PAPER MILLS LTD	C28	USEO-C	UNIV SEOUL IND COOP GROUP
C14	OSAG-C	OSAKA GAS CO LTD	C29	HITF-C	HITACHI ZOSEN CORP
C15	UYKY-N	UNIV KYUSHU	C30	UYKR-C	UNIV KOREA RES & BUSINESS FOUND

C24, C27, and C29. In the *electrolyte* subdomain, the following organisations do not have solutions: C4, C6, C13, C17, C23, and C26. Details are shown in Table 4. This indicates that these organisations still lack effective solutions to improve the photoelectric conversion efficiency of technology components or that they have not yet carried out relevant research. Figures 3–5 show our organisation correlation map based on *S* structures; the size of the nodes represents the number of related SAO structures. The red nodes represent the first patent assignee, and the green nodes represent the second patent assignee.

Potential R&D partners' identification

As can be seen in Figure 3[1], many organisations have a strong connection in research in the *photo anode* subdomain. Dainippon Printing Co. Ltd. (C7) has the most solutions, with 22 SAO structures. Fujikura Ltd. (C3) and Samsung SDI Co. Ltd. (C5) have the second most solutions, with 21 SAO structures each. However, their solutions seem distinct. C3 has solutions such as electrode substrates, metal wiring layers, and electrically conductive films. C5 has solutions such as transparent conductive films and carbon nanotubes. Organisations can look at the information we provide and decide whether the firms are potential partners. From the Literature Review, we know that Toyota Chuo Kenkyusho KK (C10) has a close cooperation with Aisin Seiki KK (C12). Also, we can see from the map that they have almost the same solutions. Semiconductor electrodes, transparent electrodes, and light-scattering layers are their solutions to the efficiency problem; therefore, we can foresee further cooperation in R&D innovation between these two firms. Konica Corp (C1), which also has similar solutions, could be considered as their potential partner. However, Nat. Inst. Advanced Ind. Sci. & Technology (C8) and Sekisui Chem. Ind. Co. Ltd. (C20) have relatively weak similarities. From the extracted

Table 3. Solution examples extracted on DSSCs.

Solution name	Solution description
Semiconductor electrode	A photoelectrode consists of a semiconductor electrode, whose light receiving surface connects to a transparent electrode
Metal complex	Metal complex dye comprises metal or metal ion, ligand (I) and ligand (II). The ligand (I) has at least 2 heterocyclic rings (I) containing the nitrogen atom, which coordinates to metal or metal ion, and has at least one group, which dehydrates hydroxyl groups contained on the metal-oxide surface
Iodine compound	An electrolyte solution that contains iodine has the iodine compound as an iodine ion source and the benzimidazole derivative with 3-11C saturated hydrocarbon group directly bonded to the benzimidazole ring

Table 4. Distribution of organisations without solutions.

Subdomain	Organisations without solutions
Photo anode	None
Sensitizer	Nippon Steel Chem. Co. Ltd. (C6), Kyocera Corp. (C7), Osaka Gas Co. Ltd. (C14), Kyocera Corp. (C17), Electronics & Telecom Res. Inst. (C18), Korea Inst. Sci. & Technology (C22), Toshiba KK (C24), Irico Group Electronics Co. Ltd. (C27), and Hitachi Zosen Corp. (C29).
Electrolyte	Sharp KK (C4), Sony Corp. (C6), Mitsubishi Paper Mills Ltd. (C13), Kyocera Corp. (C17), Nippon Kayaku KK (C23), and Toppan Printing Co. Ltd. (C26)

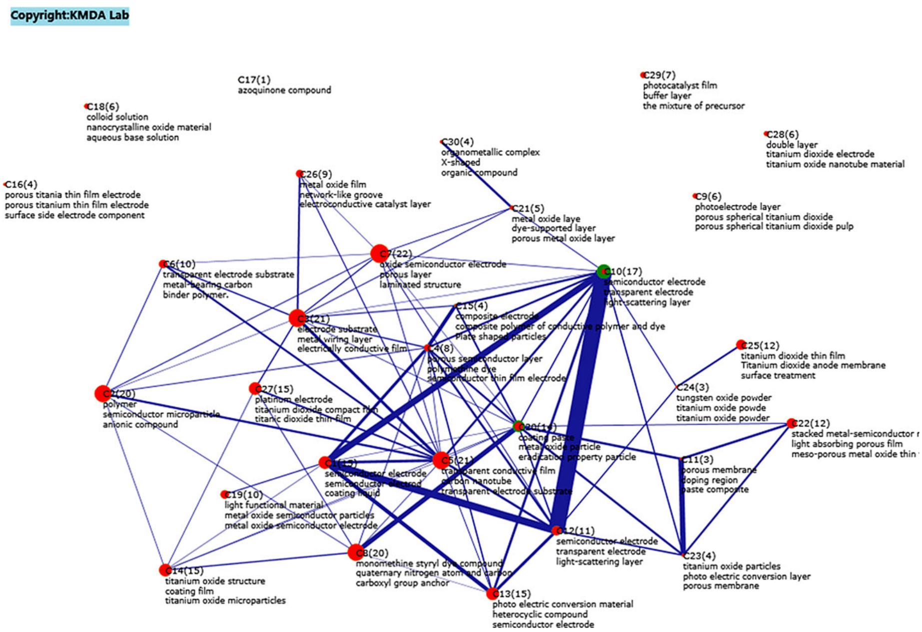

Figure 3. Organisation association analysis based on the 'S' structure of 'photo anode'.

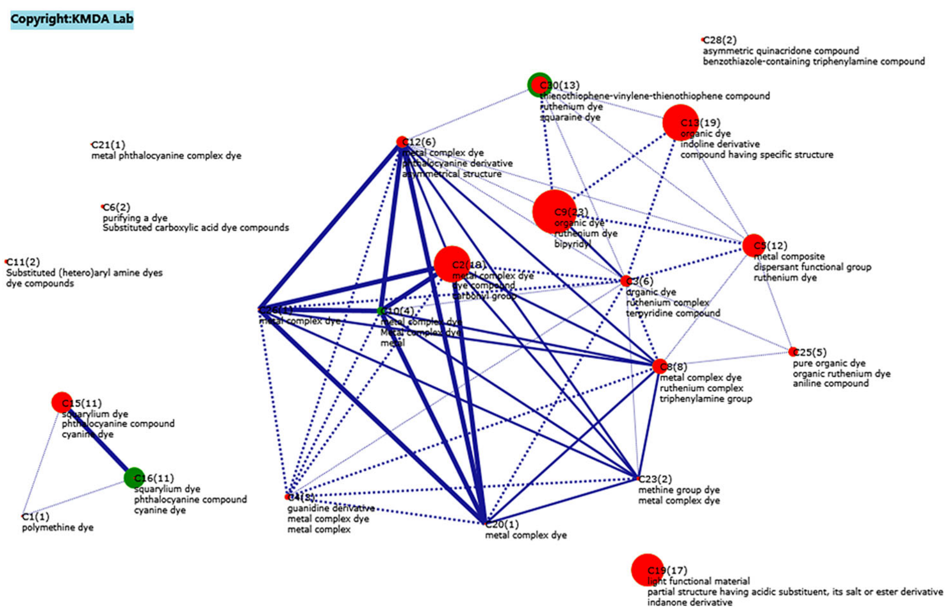

Figure 4. Organisation association analysis based on the 'S' structure of 'sensitizer'.

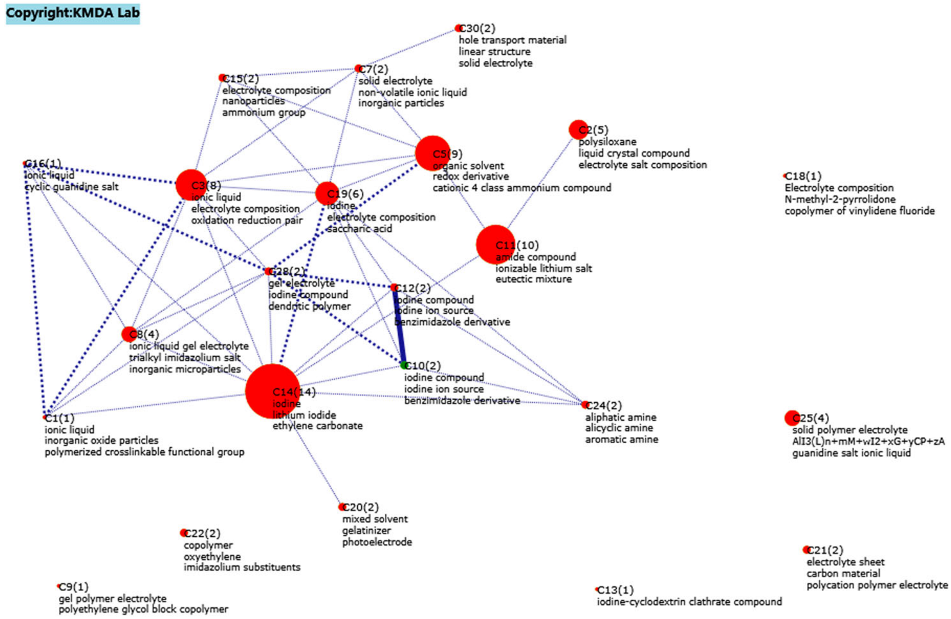

Figure 5. Organisation association analysis based on the 'S' structure of 'electrolyte'.

solutions, they show film-forming material as a common solution, but beyond that, C20 presented key solutions as coating pastes with metal-oxide particles and eradication of property particles and metal-oxide particles, whereas C8 presented key solutions as a monomethine styryl dye compound and quaternary nitrogen atoms and carbon. These organisations should consider expanding their scope for cooperation to other solutions. In addition, LG Chem. Ltd. (C11) and Nippon Kayaku KK (C23), Sharp KK (C4), and Univ. Kyushu (C15) could develop more cooperation. Other organisations can also seek their potential partners on the basis of their own research requirements and other partners' solutions. What's more, we can identify hotspot technology related to conversion efficiency in the *photo anode* subdomain, such as light-scattering layers, transparent electrodes, coating liquids, etc.

As Figure 4 shows, numerous organisations have close connections in the subdomain *sensitizer*. Dong Jin Semichem Co. Ltd. (C9), with 23 SAO structures, and Mitsubishi Paper Mills Ltd. (C13), with 19 SAO structures, have more solutions than others. They have the mutual solution of organic dye; in addition, C13 has a compound with specific structures as a solution. Based on the survey, Univ. Kyushu (C15) and Nippon Steel Chem. Co. Ltd. (C16) can strengthen their existing cooperation. Here, we offer examples of some solutions to them: squarylium dye, phthalocyanine compound, and cyanine dye. In the dense networks formed by Fuji Film Corp. (C2), Toyota Chuo Kenkyusho KK (C10), Aisin Seiki KK (C12), Sekisui Chem. Ind. Co. Ltd. (C20), and Toppan Printing Co. Ltd. (C26), the key solution 'metal complex dye' is the main solution of each organisation, but C10 has relatively limited technical strength in the *sensitizer* domain; therefore, Fuji Film Corp (C2) might consider other potential partners, such as Aisin Seiki KK (C12), Sekisui Chem. Ind. Co. Ltd. (C20), and Toppan Printing Co. Ltd. (C26). Other organisations such as Dong Jin Semichem Co. Ltd. (C9) and Fujikura Ltd. (C3) might consider developing cooperation. The hotspot technology keywords related to conversion efficiency are a metal complex dye and a ruthenium complex dye.

From the details in Figure 5, we can see the cooperative relationship among these organisations is relatively loose in the *electrolyte* subdomain. It seems that fewer organisations did R&D research on the *electrolyte* subdomain. Osaka Gas Co Ltd. (C14) has the most SAO structures in this domain. Here, we see solutions such as lithium iodide and ethylene carbonate. Other organisations could decide whether C14 can be a potential partner on the basis of their own technical solutions. The phenomenon shows that cooperation barely exists between these organisations. Toyota Chuo Kenkyusho KK (C10) and Aisin Seiki KK (C12) can strengthen their existing cooperation; their key solutions are an iodine compound and a benzimidazole derivative. Konica Corp. (C1) and Nippon Steel Chem. Co. Ltd. (C16) can develop cooperation with Fujikura Ltd. (C3) as C3 has the most solutions with eight structures; what's more, they have the same key solution of an ionic liquid. Univ. Seoul Ind. Coop Group (C28) can develop new cooperation with Samsung SDI Co. Ltd. (C5) as an organic solvent is their common solution. Other organisations can choose their desirable potential partners on the basis of their requirements.

On the whole, organisations that do not have solutions in every subdomain can consider expanding their technology range. We chose some organisations to illustrate this. If Sony Corp. (C6) wants to have a better overall range of solutions, it can choose Samsung SDI Co. Ltd. (C5) as a potential partner, as they both have a transparent electrode substrate as a solution in the *photo anode* subdomain. Other organisations can also seek potential partners in other subdomains. Irico Group Electronics Co. Ltd. (C27) does not have any solutions in the *electrolyte* and *sensitizer* subdomains. It can identify Samsung SDI Co. Ltd. (C5), Nat. Inst. Advanced Ind. Sci. & Technology (C8), and Fujikura Ltd. (C3) as potential partners. Each organisation can combine their R&D cooperation consciously and then use the information from the results to identify potential partners.

Conclusion

R&D partner selection decisions have many influencing factors. A great deal of previous work emphasises assessment of technologies. Thus, our work contributes to analyse technical solutions. This research presents a novel combination of term clumping and SAO semantic analysis, exploring the identification of R&D partners in an open innovation environment. Term clumping helps us get to the thematic dataset. Using the SAO structure in a P&S pattern, we get each organisation's solution in different dimensions. Based on the text content of solution sets, we get a correlation map of chosen research targets. The analysis indicates how different organisations adopt different solutions to specific technical problems; in addition, the proposed method can position each organisation's research focus and corporation situation, tap their potential collaboration opportunities, and help the organisations effectively identify research partners.

Further studies to extend these results can take several directions. In terms of data extraction, we only utilised abstracts of patents to extract the SAO structure. We believe that full texts could provide a more comprehensive SAO structure to reflect the intention of the articles. Moreover, it will be valuable to update the extraction tool for SAO semantic structure to reduce manual control. While this paper makes partner identification possible on the basis of the similarity between organisations' research content, we can extend the approaches from the perspective of complementarity between two research targets, certainly including characteristics such as developmental strategy, cultures of the enterprises, and R&D cooperation consciousness. Lastly, we could extend the field of application to identify partners for more organisations from different technical backgrounds.

Note

1. Here, we do not list solutions of each organization one by one; only the first three with highest frequency are listed.

Disclosure statement

There is no potential conflict of interest was reported by the authors.

Funding

This work was supported by the General Programme of the National Natural Science Foundation of China under [grant numbers 71373019, 71673024], the National High Technology Research and Development Program of China [grant number 2014AA015105].

Notes on contributors

Xuefeng Wang is a professor in the School of Management and Economics, Beijing Institute of Technology. His specialty is technology innovation management, data mining and science and technology evaluation. His current research emphasises SAO semantic analysis, technology roadmapping and forecasting innovation pathways.

Zhinan Wang is a PhD candidate in the School of Management and Economics, Beijing Institute of Technology. Her main academic research fields are technology innovation management and bibliometrics.

Ying Huang is a PhD candidate in the School of Management and Economics, Beijing Institute of Technology. His specialty is technology innovation management and technology forecasting, particularly for study of emerging technology.

Liu Yuqin is an associate professor in the School of Printing and Packaging Industrial Technology, Beijing Institute of Graphic Communication. His specialty is data mining, patent analysis and information visualisation.

Jiao Zhang is a master candidate in the School of Management and Economics, Beijing Institute of Technology. Her specialty is technology innovation management and technology evaluation.

Xiaofan Heng is a PhD candidate in the School of Management and Economics, Beijing Institute of Technology. His research interest is technology innovation management.

Donghua Zhu is a professor in the School of Management and Economics, Beijing Institute of Technology. His main academic research fields include data mining, technology innovation management, technology forecasting.

References

Ahuja, Gautam. 2000. "The Duality of Collaboration: Inducements and Opportunities in the Formation of Interfirm Linkages." *Strategic Management Journal* 21 (3): 317–343. doi:10.1002/(SICI)1097-0266(200003)21:3<317::AID-SMJ90>3.0.CO;2-B.

Angue, Katia, Cécile Ayerbe, and Liliana Mitkova. 2014. "A Method Using Two Dimensions of the Patent Classification for Measuring the Technological Proximity: An Application in Identifying a Potential R&D Partner in Biotechnology." *The Journal of Technology Transfer* 39 (5): 716–747. doi:10.1007/s10961-013-9325-8.

Baum, Joel AC, Robin Cowan, and Nicolas Jonard. 2010. "Network-independent Partner Selection and the Evolution of Innovation Networks." *Management Science* 56 (11): 2094–2110. doi:10.1080/09537325.2016.1197898.

Bergmann, Isumo, Daniel Butzke, Lothar Walter, Jens P. Fuerste, Martin G. Moehrle, and Volker A. Erdmann. 2008. "Evaluating the Risk of Patent Infringement by Means of Semantic Patent Analysis: The Case of DNA Chips." *R & D Management* 38 (5): 550–562. doi:10.1111/j.1467-9310.2008.00533.x.

Bianchi, Mattia, Alberto Cavaliere, Davide Chiaroni, Federico Frattini, and Vittorio Chiesa. 2011. "Organisational Modes for Open Innovation in the Bio-Pharmaceutical Industry: An Exploratory Analysis." *Technovation* 31 (1): 22–33. doi:10.1016/j.technovation.2010.03.002.

Brouthers, Keith D., Lance Eliot Brouthers, and Timothy J. Wilkinson. 1995. "Strategic Alliances: Choose Your Partners." *Long Range Planning* 28 (3): 2–25. doi:10.1016/0024-6301(95)00008-7.

Cao, Qing, and Qiong Wang. 2007. "Optimizing Vendor Selection in a Two-stage Outsourcing Process." *Computers & Operations Research* 34 (12): 3757–3768. doi:10.1016/j.cor.2006.01.013.

Cascini, Gaetano, Alessandro Fantechi, and Emilio Spinicci. 2004. "Natural Language Processing of Patents and Technical Documentation." In *Document Analysis Systems VI: 6th International Workshop*, edited by Simone Marinai and Andreas R. Dengel, 508–520. Berlin: Springer.

Chen, Sheu-Hua, Hong-Tau Lee, and Yi-Fen Wu. 2008. "Applying ANP Approach to Partner Selection for Strategic Alliance." *Management Decision* 46 (3–4): 449–465. doi:10.1108/00251740810863889.

Chen, Sheu Hua, Pei Wen Wang, Chien Min Chen, and Hong Tau Lee. 2010. "An Analytic Hierarchy Process Approach with Linguistic Variables for Selection of an R&D Strategic Alliance Partner." *Computers & Industrial Engineering* 58 (2): 278–287. doi:10.1016/j.cie.2009.10.006.

Chesbrough, Henry William. 2006. *Open Innovation: The New Imperative for Creating and Profiting From Technology*. Brighton, MA: Harvard Business Press.

Choi, Sungchul, Dongwoo Kang, Joohyung Lim, and Kwangsoo Kim. 2012a. "A Fact-oriented Ontological Approach to SAO-based Function Modeling of Patents for Implementing Function-based Technology Database." *Expert Systems with Applications* 39 (10): 9129–9140. doi:10.1016/j.eswa.2012.02.041.

Choi, Sungchul, Hyunseok Park, Dongwoo Kang, Jae Yeol Lee, and Kwangsoo Kim. 2012b. "An SAO-based Text Mining Approach to Building a Technology Tree for Technology Planning." *Expert Systems with Applications* 39 (13): 11443–11455. doi:10.1016/j.eswa.2012.04.014.

Cloodt, Myriam, John Hagedoorn, and Hans Van Kranenburg. 2006. "Mergers and Acquisitions: Their Effect on the Innovative Performance of Companies in High-tech Industries." *Research Policy* 35 (5): 642–654. doi:10.1016/j.respol.2006.02.007.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. 2002. "Links and Impacts: The Influence of Public Research on Industrial R&D." *Management Science* 48 (1): 1–23. doi:10.1287/mnsc.48.1.1.14273

Fabrizio, Kira R. 2009. "Absorptive Capacity and the Search for Innovation." *Research Policy* 38 (2): 255–267. doi:10.1016/j.respol.2008.10.023.

Fischer, Marco, Hendrik Jähn, and Tobias Teich. 2004. "Optimizing the Selection of Partners in Production Networks." *Robotics and Computer-integrated Manufacturing* 20 (6): 593–601. doi:10.1016/j.rcim.2004.05.007.

Fleck, James. 1997. "Contingent Knowledge and Technology Development." *Technology Analysis & Strategic Management* 9 (4): 383–398. doi:10.1080/09537329708524293.

Gambardella, Alfonso, and Anita M. McGahan. 2010. "Business-model Innovation: General Purpose Technologies and their Implications for Industry Structure." *Long Range Planning* 43 (2–3): 262–271. doi:10.1016/j.lrp.2009.07.009.

Gerken, Jan M., and Martin G. Moehrle. 2012. "A New Instrument for Technology Monitoring: Novelty in Patents Measured by Semantic Patent Analysis." *Scientometrics* 91 (3): 645–670. doi:10.1007/s11192-012-0635-7.

Guo, Junfang, Xuefeng Wang, Qianrui Li, and Donghua Zhu. 2016. "Subject–Action–Object-based Morphology Analysis for Determining the Direction of Technological Change." *Technological Forecasting and Social Change* 105 (4): 27–40. doi:10.1016/j.techfore.2016.01.028.

Hagedoorn, John. 2002. "Inter-firm R&D Partnerships: An Overview of Major Trends and Patterns Since 1960." *Research Policy* 31 (4): 477–492. doi:10.1016/S0048-7333(01)00120-2.

Hajidimitriou, Yannis A., and Andreas C. Georgiou. 2002. "A Goal Programming Model for Partner Selection Decisions in International Joint Ventures." *European Journal of Operational Research* 138 (3): 649–662. doi:10.1016/S0377-2217(01)00161-8.

Hill, Charles W. L., and Frank T. Rothaermel. 2003. "The Performance of Incumbent Firms in the Face of Radical Technological Innovation." *The Academy of Management Review* 28 (2): 257–274. doi:10.2307/30040712.

Huang, Ying, Jannik Schuehle, Alan L. Porter, and Jan Youtie. 2015. "A Systematic Method to Create Search Strategies for Emerging Technologies Based on the Web of Science: Illustrated for 'Big Data'." *Scientometrics* 105 (3): 2005–2022. doi:10.1007/s11192-015-1638-y.

Huang, Ying, Yi Zhang, Jing Ma, Alan L. Porter, Xuefeng Wang, and Ying Guo. 2016. "Generating Competitive Technical Intelligence Using Topical Analysis, Patent Citation Analysis, and Term Clumping Analysis." In *Anticipating Future Innovation Pathways Through Large Data Analysis*, edited by Tugrul U. Daim, Denise Chiavetta, Alan L. Porter and Ozcan Saritas, 153–172. Cham: Springer International.

Kang, Ki H., and Jina Kang. 2010. "Does Partner Type Matter in R&D Collaboration for Product Innovation?" *Technology Analysis & Strategic Management* 22 (8): 945–959. doi:10.1080/09537325.2010.520473.

Keil, Thomas, Markku Maula, Henri Schildt, and Shaker A. Zahra. 2008. "The Effect of Governance Modes and Relatedness of External Business Development Activities on Innovative Performance." *Strategic Management Journal* 29 (8): 895–907. doi:10.1002/smj.672.

Kim, Youngho, Yingshi Tian, Yoonjae Jeong, Ryu Jihee, and Sung-Hyon Myaeng. 2009. "Automatic Discovery of Technology Trends from Patent Text." In *Proceedings of the 2009 ACM Symposium on Applied Computing*, 1480–1487. Honolulu: ACM.

Lai, Kuei-Kuei, and Shiao-Jun Wu. 2005. "Using the Patent Co-citation Approach to Establish a New Patent Classification System." *Information Processing & Management* 41 (2): 313–330. doi:10.1016/j.ipm.2003.11.004.

Lane, Peter J., and Michael Lubatkin. 1998. "Relative Absorptive Capacity and Inter-organizational Learning." *Strategic Management Journal* 19 (5): 461–477. doi:10.1002/(SICI)1097-0266(199805)19:5<461::AID-SMJ953>3.0.CO;2-L.

Laursen, Keld, and Ammon Salter. 2006. "Open for Innovation: The Role of Openness in Explaining Innovation Performance among U.K. Manufacturing Firms." *Strategic Management Journal* 27 (2): 131–150. doi:10.1002/smj.507.

López, Alberto. 2008. "Determinants of R&D Cooperation: Evidence from Spanish Manufacturing Firms." *International Journal of Industrial Organization* 26 (1): 113–136. doi:10.1016/j.ijindorg.2006.09.006.

Mikhailov, L. 2002. "Fuzzy Analytical Approach to Partnership Selection in Formation of Virtual Enterprises." *Omega* 30 (5): 393–401. doi:10.1016/S0305-0483(02)00052-X.

Moehrle, Martin G., Lothar Walter, Anja Geritz, and Sandra Müller. 2005. "Patent-based Inventor Profiles as a Basis for Human Resource Decisions in Research and Development." *R&D Management* 35 (5): 513–524. doi:10.1111/j.1467-9310.2005.00408.x.

Mowery, David C., Joanne E. Oxley, and Brian S. Silverman. 1996. "Strategic Alliances and Interfirm Knowledge Transfer." *Strategic Management Journal* 17 (S2): 77–91. doi:10.1002/smj.4250171108.

Mowery, D. C., J. E. Oxley, and B. S. Silverman. 1998. "Technological Overlap and Interfirm Cooperation: Implications for the Resource-based View of the Firm." *Research Policy* 27 (5): 507–523. doi:10.1016/S0048-7333(98)00066-3.

Nijssen, Edwin J., Rik Van Reekum, and Henriëtte E. Hulshoff. 2001. "Gathering and Using Information for the Selection of Technology Partners." *Technological Forecasting and Social Change* 67 (2–3): 221–237. doi:10.1016/S0040-1625(00)00097-4.

Noseleit, Florian, and Pedro de Faria. 2013. "Complementarities of Internal R&D and Alliances with Different Partner Types." *Journal of Business Research* 66 (10): 2000–2006. doi:10.1016/j.jbusres.2013.02.025.

O'Regan, Brian, and Michael Gratzel. 1991. "A Low-cost, High-efficiency Solar Cell Based on Dye-sensitized Colloidal TiO2 Films." *Nature* 353 (6346): 737–740. doi:10.1038/353737a0.

Park, Inchae, Yujin Jeong, Byungun Yoon, and Letizia Mortara. 2015. "Exploring Potential R&D Collaboration Partners Through Patent Analysis Based on Bibliographic Coupling and Latent Semantic Analysis." *Technology Analysis & Strategic Management* 27 (7): 759–781. doi:10.1080/09537325.2014.971004.

Porter, Alan L., and Scott W. Cunningham. 2004. *Tech Mining: Exploiting New Technologies for Competitive Advantage*. Hoboken, NJ: Wiley.

Samadhi, T. M. A. Ari, and K. Hoang. 1998. "Partners Selection in a Shared-CIM System." *International Journal of Computer Integrated Manufacturing* 11 (2): 173–182. doi:10.1080/095119298130903.

Sarkis, Joseph, Srinivas Talluri, and A. Gunasekaran. 2007. "A Strategic Model for Agile Virtual Enterprise Partner Selection." *International Journal of Operations & Production Management* 27 (11): 1213–1234. doi:10.1108/01443570710830601.

Solesvik, Marina Z., and Sylvia Encheva. 2010. "Partner Selection for Interfirm Collaboration in Ship Design." *Industrial Management & Data Systems* 110 (5): 701–717. doi:10.1108/02635571011044731.

Solesvik, Marina Z., and Paul Westhead. 2010. "Partner Selection for Strategic Alliances: Case Study Insights From the Maritime Industry." *Industrial Management & Data Systems* 110 (6): 841–860. doi:10.1108/02635571011055081.

Talluri, Srinivas, R. C. Baker, and Joseph Sarkis. 1999. "A Framework for Designing Efficient Value Chain Networks." *International Journal of Production Economics* 62 (1–2): 133–144. doi:10.1016/S0925-5273(98)00225-4.

Teddlie, Charles, and Abbas Tashakkori. 2009. *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Thousand Oaks, CA: Sage.

Verbitsky, Mikhail. 2004. "Semantic TRIZ." Accessed September 10. https://triz-journal.com/semantic-triz/.

Vicente-Gomila, Jose M. 2014. "The Contribution of Syntactic–Semantic Approach to the Search for Complementary Literatures for Scientific or Technical Discovery." *Scientometrics* 100 (3): 659–673. doi:10.1007/s11192-014-1299-2.

Wang, Ming-Yeu. 2012. "Exploring Potential R&D Collaborators with Complementary Technologies: The Case of Biosensors." *Technological Forecasting and Social Change* 79 (5): 862–874. doi:10.1016/j.techfore.2011.11.002.

Wang, Tien-Chin, and Yueh-Hsiang Chen. 2007. "Applying Consistent Fuzzy Preference Relations to Partnership Selection." *Omega* 35 (4): 384–388. doi:10.1016/j.omega.2005.07.007.

Wang, Xuefeng, Rongrong Li, Shiming Ren, Donghua Zhu, Meng Huang, and Pengjun Qiu. 2014. "Collaboration Network and Pattern Analysis: Case Study of Dye-sensitized Solar Cells." *Scientometrics* 98 (3): 1745–1762. doi:10.1007/s11192-013-1180-8.

Wang, Xuefeng, Pengjun Qiu, Donghua Zhu, Liliana Mitkova, Ming Lei, and Alan L. Porter. 2015. "Identification of Technology Development Trends Based on Subject–Action–Object Analysis: The Case of Dye-sensitized Solar Cells." *Technological Forecasting and Social Change* 98: 24–46. doi:10.1016/j.techfore.2015.05.014.

Wu, Wann Yih, Hsi-An Shih, and Hui-Chun Chan. 2009. "The Analytic Network Process for Partner Selection Criteria in Strategic Alliances." *Expert Systems with Applications* 36 (3): 4646–4653. doi:10.1016/j.eswa.2008.06.049.

Yoon, Janghyeok, Hyunseok Park, and Kwangsoo Kim. 2013. "Identifying Technological Competition Trends for R&D Planning Using Dynamic Patent Maps: SAO-based Content Analysis." *Scientometrics* 94 (1): 313–331. doi:10.1007/s11192-012-0830-6.

Yoon, Byungun, and Bomi Song. 2014. "A Systematic Approach of Partner Selection for Open Innovation." *Industrial Management & Data Systems* 114 (7): 1068–1093. doi:10.1108/IMDS-03-2014-0086.

Zhang, Yi, Alan L. Porter, Zhengyin Hu, Ying Guo, and Nils C. Newman. 2014a. "'Term Clumping' for Technical Intelligence: A Case Study on Dye-sensitized Solar Cells." *Technological Forecasting and Social Change* 85 (6): 26–39. doi:10.1016/j.techfore.2013.12.019.

Zhang, Yi, Xiao Zhou, Alan L. Porter, and Jose M. Vicente Gomila. 2014b. "How to Combine Term Clumping and Technology Roadmapping for Newly Emerging Science & Technology Competitive Intelligence: 'Problem & Solution' Pattern Based Semantic TRIZ Tool and Case Study." *Scientometrics* 101 (2): 1375–1389. doi:10.1007/s11192-014-1262-2.