Scientometrics (2022) 127:5519–5550
https://doi.org/10.1007/s11192-022-04306-9

Choosing the right collaboration partner for innovation: a framework based on topic analysis and link prediction

Yan Qi[1] · **Xin Zhang**[2] · **Zhengyin Hu**[2] · **Bin Xiang**[2] · **Ran Zhang**[1] · **Shu Fang**[2]

Received: 11 May 2021 / Accepted: 11 February 2022 / Published online: 12 March 2022

Abstract

Selecting the right collaboration partner is one of the most important contributors to success in collaborative innovation. Accordingly, numerous methods for selecting an appropriate partner have been developed to guide would-be collaborators in their search. Most rely on bibliographic information, which may be easier for that data is readily available and relatively normalized. However, with the benefit of today's text mining and fusion techniques, it is possible to mine the content of papers and patents so as to result in far more nuanced and advantageous choices. In this article, we explore how to select partners for collaborative innovation by combining the characteristics of the authors of paper and patent documents as well as their content. Drawing on existing research, we developed a systematic framework that relies on topic analysis and link prediction. With a corpus of papers and patents assembled, the framework extracts correlated scientific and technological topics followed by a list of author institutions and a list of patentees. These organisations are parsed and evaluated using two indicators of innovation—capability and openness—to result in two separate ranked lists. Two integrated collaboration networks that include both author institutions and patentees are then built, and a link prediction method identifies missing links with a high likelihood of fruitful cooperation. A case study on hepatitis C virus research shows that the ranking procedure and the link prediction method can be used either together or separately to effectively identify collaborative innovation partners. Our results provide significant quantitative evidence for policymakers who are looking to foster cooperation between research institutions and/or high-tech enterprises. Our research may also serve as the basis for further in-depth research on collaborative innovation, R&D cooperation, and link prediction theories and methods.

Keywords Collaborative innovation · Partner selection · Topic analysis · Link prediction

✉ Yan Qi
qi.yan@imicams.ac.cn

✉ Xin Zhang
zhangxin@clas.ac.cn

[1] Institute of Medical Information/Medical Library, Chinese Academy of Medical Sciences/Peking Union Medical College (CAMS&PUMC), No. 3, Yabao Road, Beijing 100020, China

[2] Chengdu Library and Information Center, Chinese Academy of Sciences, No. 16, South Section 2, Yihuan Road, Chengdu 610041, China

Introduction

With the increasing rate of economic globalisation, open innovation has become more vital than ever. "Continuous innovation", i.e., the "modest, incremental, ongoing upgrades or enhancements of existing technologies, services or products" (University of California, 2015), and sustainable development are among two highly influential factors that allow enterprises to compete in today's fiercely competitive and complex global markets (Kang et al., 2019). However, as technological innovation grows in its complexity, it is becoming more and more difficult for one organisation to achieve innovation through individual efforts alone. This is because it is almost impossible for most institutions to master all the knowledge and skills required to take an idea from its conception right through to market (Gnyawali & Park, 2011). As such, many organisations are turning to open cooperation with other players in the innovation space—to conduct R&D, to acquire new and complementary knowledge and resources, to relegate the work required into spheres of expertise, and so on (Chesbrough, 2003; Wang et al., 2015). Universities and research institutions, in particular, are also seeking collaborations to increase the pace of their scientific and technological innovation, which is becoming more and more difficult. Thus, from a macro perspective, there is a need to build open collaborative innovation among the various figures involved in scientific and technological knowledge creation, innovation, and application (Chen & Yang, 2012; Hagedoorn, 2002). In other words, in today's dynamic and increasingly complex world, fostering innovation ecosystems is essential for technological development.

In this context, many strategies to promote collaborative innovation have attracted attention. These include concepts such as enterprise alliances and industrial clusters, along with the many national, regional, discipline-level, or field-level science parks and collaborative innovation centres that have been established. Further, policies to promote a range of cooperation innovations have also realised significant achievements, such as (1) inter-enterprise cooperations, (2) industry-university cooperations, (3) industry-university-research synergies, and so forth (Wang & Huang, 2020; Wen & Kobayashi, 2001). These partnerships have not only advanced science and technology but also generated profits from the creation of new products.

However, there have been several problems and challenges, such as cooperation between some enterprises can lead to the problem of competition in the same market. In some cases, this has resulted in mistrust that weakens cooperative ties (Ji et al., 2020). One-size-fits-all policies and homogeneous approaches to technological development can also make it difficult for alliances to differentiate their solutions in the marketplace. In turn, this can promote vicious competition. The correlations between innovation entities and industries might also be too weak to form an industrial symbiosis (Huang, 2010; Wang et al., 2014, 2016). Thus, initiatives fail to achieve the expected synergies (Li et al., 2011). What can also contribute to failed alliances is the inherent conflict resulting from divergent goals, opportunistic behaviours by one's partners, and cultural differences (Doz, 1996; Kale et al., 2000). And a failed alliance is usually very costly to all parties concerned, both financially and reputationally.

Many researchers and industry practitioners alike have suggested that a judicious process for selecting partners is one of the most important factors in a successful collaboration (Dacin et al., 1997; Geum et al., 2013; Ireland et al., 2002). But how does one identify and select the right partner? The existing research on methods for identifying partners can be divided into several categories. The first one is Surveys. Some have used surveys to develop guidelines for selecting the right partners (Carayannis et al., 2000). These surveys are designed to interrogate the various characteristics that should be considered in a good partnership, such as compatible goals and complementary skills (Brouthers et al., 1995), knowledge of the environment, reputation (Ariño et al., 1997), financial assets (Hitt et al., 2000), prior experience (Nielsen, 2003), technological relatedness and prior ties (Petruzzelli, 2011), geographic proximity (Capaldo & Petruzzelli, 2014), culture (Elia et al., 2019), firm age (Ardito et al., 2019a), motivation (Rajalo & Vadi, 2021), and so on. These research results have provided a wealth of evaluation indicators for the partner selection process. However, some of these indicators rely on subjective judgments rather than quantitative factors, which has certain limitations.

Quantitative and literature-based methods have also been used. Nguyen (2021), for example, applied modern statistical techniques and the Grey method to help businesses select appropriate partners in a supply chain. Most other studies have employed bibliometric techniques on a corpus of patents or publications (rarely both). In fact, bibliometric analysis has been widely employed to extract meaningful knowledge from the scientific literature to support expert decision-making. In these pursuits, particular focus is often given to commercially important patents. These are the literature-based methods, and they generally follow one of four broad methodologies: indicators; network analysis and link prediction; text mining; or fusion methods.

Indicator methods seek to investigate specific aspects of a corpus and so commensurate indicators are designed to measure the desired dimensions. Researchers generally develop one or several mathematical models and then use those models to measure characteristics of the corpus—for example, Geum et al. (2013) designed 14 indices based on the bibliometric and citation information of scientific publications and patents to search for and assess suitable R&D partners.

Methods based on network analysis and link prediction focus on historical partnerships and leverage network topologies for predictions. Examples include Yan & Guns (2014), who constructed Author-, institution-, and country-level collaboration networks to examine, predict, and recommend collaborations. Focusing on China, Chen et al. (2021) explored partner selection in interorganisational patent cooperation. With this methodology, some researchers have also fused indicators into the mix, such as network 'neighbours' and 'paths' (Yu et al., 2017), the similarity of network paths and the similarity of research interests (Liu & Sun, 2017), etc.

Text mining methods attempt to extract meaningful insights from a corpus. Here, scholars pay attention to the content in scientific and technological literature, using text mining technology to identify partners based on the understandings extracted. For instance, Jeon et al. (2011) mined the interrelationships between words in patent claims to identify potential partners with the desired technology. Wang et al. (2015, 2017) proposed novel processes for identifying R&D partners on the basis of goal or solution similarities through subject-action-object semantic analysis. Kang et al. (2019) further suggested a methodology of partner selection for sustainable industry-university cooperation based on latent Dirichlet allocation (LDA) topic modelling.

Fusion methods consider both the internal characteristics of a document's contents and its external characteristics, such as its co-authors, attributions, and/or references. In this category, Park et al. (2015) explored potential R&D collaboration partners through patent analysis based on bibliographic coupling and latent semantic analysis. Alternatively, Ding & Guo (2021) developed a method of mining potential cooperative relationships. The method involves determining how similar the content of two authors' articles are, and examining cooperative network structures.

While each of the above studies address important research questions, this study extends previous findings in several ways. First, as we will show, we have expanded the theoretical foundations of these methods by examining both patents and papers and by considering both the internal and external characteristics of both types of documents. Most existing studies rely solely on patent documents due to better data accessibility and their importance in the marketplace (Geum et al., 2013). Even when studying industry-university-research cooperations, researchers like Kang et al. (2019) mostly use patent data. While Geum et al. (2013) used both patents and papers as data sources, the relevant indicators and data were measured separately. Moreover, only the external features of the literature were used; the content of the literature was not considered as is the case with text mining methods. However, it should be noted that there are increasingly many connections or linkages between science and technology, which can be approximately represented by papers and patents (Lo, 2009; Narin et al., 1997; Verbeek et al., 2002). In some disciplines or fields, the knowledge flows between science and technology are strong, given the large number of citations between patents and papers (Qi, 2019; Schmoch, 1997; Yin, 2012). Moreover, some researchers regard the cooperative R&D activities of public scientific research institutes and enterprises as a kind of linkage between scientific and technological knowledge (Garg, 2001). We believe that cooperation in the context of science and technology association is necessary, and that this kind of complementary cooperation between different types of knowledge and skills is relatively more strategic and sustainable. The result should be better collaborative outcomes for innovation while avoiding the vicious competition often associated with homogeneity—at least to some extent. However, to our best knowledge, no study on selecting partners for collaborative innovation projects has yet integrated papers and patents while also examining both the content of the documents as well as their meta data. This study provides a method for doing just that. Importantly, the method fuses both paper- and patent-based collaboration information into a single network to enable link prediction.

Second, the method, uniquely, does not divide universities from enterprises nor do we only recommend partnerships between the two. The framework may also recommend university-university pairs or enterprise-enterprise pairs. The only consideration for recommending a partner is knowledge of the literature. This is somewhat different from existing studies, which either match up universities with enterprises or enterprises with other enterprises. We opted not to take this approach on several counts. Increasingly, many universities and research institutes are applying for patents or establishing school-run enterprises, such as innovation incubation centers. Etzkowitz et al. (2000) refers to this as the rise of the "entrepreneurial university". Therefore university-university partnerships can be a valid and even sought-after option. Further, enterprises are also increasingly participating in the publication of papers—both alone and as co-authors with university scholars. Thus, we argue that our method might yield better partner options for real-world scenarios if based solely on a knowledge-based approach rather than an institution-based approach.

In summary, this article builds on the theories and methods of existing research to produce a framework for selecting innovation partners that relies on both the content and the meta data of papers and patents. Drawing on existing research by Geum et al. (2013); Park et al. (2015); Kang et al. (2019); Han et al. (2021); Cui et al. (2021), we have developed a systematic methodology comprising text mining, topic analysis, network analysis, and link prediction. Topics are extracted from the document contents via a text mining technique and then analysed for similarity to determine compatible areas of research or development between organisations (Cui et al., 2021). Network analysis and link prediction are then used to predict the probability of future cooperation based on historical relationships (Yan

& Guns, 2014). A case study on the hepatitis C virus shows this method achieves better results than the existing fusion methods.

We hope that this study can serve as the basis for further in-depth research on the theories and methods associated with the search for collaborative innovation partners. We also hope to provide references for the expansion and practice of link prediction methods. For science & technology managers and innovation entities, such as enterprises, universities, and scientific research institutes, we hope to provide a useful tool and decision-making reference for selecting the right innovation partner for the right project.

Literature review and research framework

Selecting and identifying innovation partners

The rapid development of science and technology has led to the accelerated upgrading of products and services, while the complexity of technological innovation and product innovation has urged enterprises to break through organisational limitations and cooperate with other enterprises or research institutions to carry out open innovation (Ji et al., 2020). This wave of open innovation has given birth to various forms of cooperation, such as technology alliances, strategic alliances, and innovation alliances. R&D cooperation or collaborative innovation with external partners allows firms to access external resources, reduce the time it takes to innovate, share the associated risks, and improve overall performance (Alexiev et al., 2016; Ardito et al., 2019b; Galati et al., 2019; Clauss & Kesting, 2017; Hagedoorn, 1993; Laursen & Salter, 2006; Leiponen & Helfat, 2010). Enterprise-university or industry-university-research alliances have also received a much attention. As Friesike et al. (2015) point out, a symbiotic relationship between private and public institutions makes sense given their different backgrounds. Research institutions can enable research capabilities, and private companies can contribute know-how when it comes to commercialisation. Public institutions have also changed to foster more collaborative and open innovation (Jugend et al., 2020), while the increasing collaborations between businesses and universities have led to new insights and innovations (Ottonicar et al., 2020).

It is well known that choosing the right partner is a key factor affecting the success of collaborative innovation and alliance performance. In fact, choosing an inappropriate partner could turn out to be even more costly and risky than trying to go it alone (Brouthers et al., 1995). As such, identifying and selecting partners has long been a focus for both academia and industry (Geum et al., 2013; Ireland et al., 2002). There has been a considerable amount of literature investigating partner characteristics as determinants of inter-firm alliances—characteristics like firm size, R&D intensity, and sales volume (Becker & Dietz, 2004; Veugelers, 1997), inter-firm trust and mutual interests (Hagedoorn et al., 2008; Inkpen & Currall, 1997), cultural fit (Elia et al., 2019; Littler et al., 1995), geographic and organisational proximity (Capaldo & Petruzzelli, 2014), and resource fit (Barney, 1991; Rothaermel, 2001).

As mentioned in the Introduction, existing partner identification and selection methods can be roughly divided into several categories: index system methods that require subjective judgment, quantitative model methods, and literature-based methods, with the literature-based methods further divided into several subcategories. In terms of index-based methods, many scholars have tried to establish guidelines for the selection of potential partners. Nielsen (2003), for example, identified seven criteria (with 21 sub-criteria) for partner selection: technological expertise, marketing systems and status, local operational expertise, competitive strength, production efficiency, positive prior experience, and labor negotiation expertise. Liu & Wang (2005) select the best partner based on cultural syncretism, complementarity, and credit status. Wu et al. (2009) proposes an integrated approach of analytic network process (ANP) to consider both tangible and intangible factors and to optimize the paid off earn by company from strategic alliance. The five factors ranked from highest to lowest importance are complimentary capabilities, intangible assets, marketing knowledge capability, degree of fitness and characteristics of partners. Tian (2014) selects strategic alliance partners for high-tech enterprises using strategic synergy, cultural compatibility, strength matching, and complementary resources as indicators. Focusing on the process of cooperation, Cao and Song (2016) analysed the "demand" of an enterprise in different stages of cooperative R&D and built a "multi-stage & multi-demand" index system for selecting industry-university-research cooperation partners.

In terms of quantitative studies, we found research that integrates quantitative evaluation indices into quantitative models, such as Zhang et al. (2015) who built an innovation ability evaluation index and then applied a bilateral matching method to an entity-matching matrix combined with a weighted index to get the best matching scheme. Wang (2016) constructed an evaluation index system for industry-university-research partner selection for collaborations involving the internet of things (IoT). The index system considered the innovation environment, innovation input capacity, and innovation output capacity, and the subsequent three-stage mathematical model involved analytic network process (ANP), data envelopment analysis (DEA) and Grey relational analysis (GRA) to make partner selection more objective, effective, and comprehensive. More recently, some studies have treated partner selection as a multi-objective optimisation problem. As an example, Ionescu & Vernic (2021) proposed a new approach for dealing with unfeasible task scheduling solutions—the multi-objective symbiotic organisms search for scheduling (MOSOSS) algorithm, which is based on the multiple objective symbiotic organism search (MOSOS) algorithm (Tran et al., 2016)—and adapted it for a rather general formulation of the partner selection problem.

Among the literature-based methods, the first subcategory is the index method based on bibliographic information. In this vein, Geum et al. (2013) designed a systematic framework containing 14 indicators from four dimensions including technology strength, R&D openness, R&D linkage, and collaboration effects. The result is a partner selection method based on the classified statistical data of scientific papers and patents. Song et al. (2016) proposed a method for selecting potential R&D partners based on several criteria: the value of each firm's patent portfolio; the degree to which the firms' resources match; and the Shapley value, which assesses the contribution of each potential partner to a combined patent portfolio.

The second subcategory is the network analysis/link prediction methods using single or multiple network indicators, or in combination with other external feature indicators based on bibliographic data. Here, Liben-Nowell & Kleinberg (2007) put forward the problem of link prediction in social networks, providing a definition of similarity based on network structures. They subsequently applied two types of indices based on network nodes and network paths to inform link prediction. Notably, most studies published since the Liben-Nowell & Kleinberg (2007) article have focused on improving those indicators. The string of link prediction indices based on similarity that have emerged as a result include: the common neighbour index, cosine similarity, the Jaccard index, Adamic-Adar, resource allocation, the Leicht-Holme-Newman index, the local path index, the Katz index, random walk with restart model, and others. Notable studies include Yu et al. (2017), who established a new collaboration recommendation model based on scientific collaboration network 'neighbours' and 'paths' and then conducted an empirical study to examine the model at the individual, institution, and regional levels. Guns & Rousseau (2014) built a weighted cooperation network in the research field of malaria and tuberculosis and conducted collaboration prediction and recommendation. Yan & Guns (2014) analysed the papers of 59 periodicals in the field of library and information in an empirical study and compared the predicted results under various indicators. Wang et al. (2019) proposed two indicators: IDF (Institutional-Document Frequency) and ICCR (Institutional Cumulative Cooperation Ratio) to reflect the similarity of the authors' preferences in choosing cooperating institutions. They also combined these two new indicators with four commonly-used indicators being the common neighbour index, Adamic-Adar, the local path index, and the Katz index. They constructed eight weighted algorithms to predict potential scientific collaborations and found that weighted prediction algorithms based on fusion indicators yielded better prediction results. Chen et al. (2021) also used eight link prediction approaches commonly used in social networks to explore the partner selection in China's interorganisational patent cooperation network. Furthermore, other methods were developed for cooperative prediction based on maximum likelihood estimation, probability graph models, and so on. In this study, we use a link prediction method from existing research and incorporate it into a fusion network of author and patentee organisations to predict potential collaborations.

The third subcategory is methods that use natural language processing (NLP) techniques to analyse the "intrinsic" content of the literature. Compared with the previous two subcategories, this method of text analysis pays more attention to the content of R&D cooperation and can more fully explore the role of literature as a technical intelligence resource for identifying partners. For example, Jeon et al. (2011) used a patent text mining technology to select suitable R&D partners based on the idea of matching specific technical needs. Wang et al. (2015) used subject-action-object (SAO) text mining technology to identify institutions with similar R&D goals from the scientific literature as potential R&D partners. Kang et al. (2019) suggest a systematic methodology that combines latent Dirichlet allocation (LDA) topic modelling and a clustering algorithm to identify the best college partners. Their framework classifies the subcategories of a particular technological domain, identifying potential partners in each subcategory. Wang et al. (2021) improved the method of Makri et al. (2010) and provided a framework for exploring the technology complementarity between enterprises in a quantitative manner with a hierarchical latent Dirichlet allocation (hLDA) topic model. Their system involves text-mining patent data and using the intelligence extracted to identify R&D opportunities, find appropriate acquisition targets, and recommend potential collaborators.

In view of the advantages of each of these broad methodologies, scholars have also explored combining these methods into a fusion that selected partners based on the intrinsic content of the literature and external bibliographic information. Common criteria for selection includes similar research content and the topological characteristics of cooperation networks. Research in this subcategory includes the fusion of the previously mentioned text-based methods with index-based methods and network-based methods. For example, Park et al. (2015) proposed a framework that combines two analytic methods of measuring technological similarity—bibliographic coupling analysis based on citation relationships, i.e., the bibliographic information in patents, and latent semantic analysis based on semantic similarity, i.e., the textual information in patents. In the research of Song et al. (2016), specific technological needs are represented as action-object (AO) structures. The relevance of the patents to an organisation's technological needs is then measured as the normalised frequency of co-occurrence between the A and O elements at the sentence level. In turn, this information is used to identify potential R&D partners. Further, technology share and R&D emphasis are used as indicators to reflect a firm's patent portfolio. The Shapley value is then used to assess the contribution of each potential partner to the value of an integrated patent portfolio.

Xu et al. (2016) used the relatively advanced 3-mode network analysis method to consider the correlations between technical subjects in patent literature. Ding & Guo (2021) integrated a path similarity index within a complex network with a certain weight and constructed a potential cooperative relationship mining method based on the similarity of the author's research content and the cooperative network structure. Han et al. (2021) suggest that a network's nature and structure are the basis for discovering potential cooperative links, and that knowledge attributes are an important factor in the cooperative relationships. Thus, they not only define the problem of knowledge-cooperative network link prediction but also introduce knowledge attribute indicators as a basis for knowledge-cooperative networks. Cui et al. (2021) use LDA topic modelling to mine the text of patents for candidate partners, dividing them into different technical topics. A patent evaluation system that considers both professional and collaborative capabilities then determine the best R&D partners under each technology theme.

Drawing on existing research, this study also employs a fusion method that combines content mining, topic analysis, network analysis, and link prediction to identify and select collaborative innovation partners.

The correlations between science and technology

Reconsidering the themes of science and technology relationship research, De Solla Price (1965) highlighted that knowledge may flow from science to technology and vice versa, even if each has a unique knowledge accumulation structure. Bhattacharya et al. (2003) argue that scientific research provides a source of knowledge for technological progress, and that new technological inventions can, in turn, inform scientific research, inspire ideas, and expand novel research spaces. Moreover, these scholars contend that both evolve through knowledge transfer and feedback. Ju & Liu (2004) also confirm a two-way relationship between science and technology. From one perspective, science offers a theoretical basis for technology; from another perspective, the development of science requires technology that can test the accuracy of scientific theories. That is, science research and technological improvement focus on different categories of innovation in the innovation chain. Given that they are simultaneously independent and interrelated, we believe that there is not only room for cooperation where they meet but, additionally, this type of collaboration can solve a problem from both a scientific and a technical perspective or from different stages, thus avoiding the problem of intense competition.

Many scholars have conducted in-depth research on the relationship between science-technology from different perspectives (Garfield, 1984; Meyer, 2000; Narin et al., 1997; Verbeek et al., 2003), and many studies have been undertaken to uncover the links between scientific and technical knowledge. For example, Garg (2001) looked at the cooperative R&D activities of public scientific research institutes and enterprises. Autant (2001) examined the geographic association between the private and public sectors. Maraut & Martínez (2014) believe author–inventors are the key researchers at the centre of science–industry linkages, and Wang & Guan (2011) also use author–inventors to measure science-technology interactions. The most common method of revealing these links is co-occurrence analysis or citation analysis, usually of either academic papers or patents, or much more rarely both. Papers are typically regarded as manifestations of scientific research achievements, while patents are the outcome of technological innovation. Based on the number or content characteristics of the papers and patents, this method can reveal the linkage between science and technology quantitatively and microscopically. For instance, Carpenter & Narin (1983) pioneered the use of a patent's essay citations to measure the interaction between science and technology, while the CHI Corporation established the scientific linkage index of technology. Since then, many other indices have been established.

Other studies have focused on literature content, such as Bassecoulard & Zitt (2004), who used the Chemical Abstracts Database as a data source to discover science and technology associations by establishing a vocabulary correspondence table between papers and patents. Sun & Ding (2018) judged the relationships between science and technology in a specific field by discovering the similarities and differences of knowledge genes in paper and patent collections. Xu et al. (2019); Liu et al. (2019); Wang (2020) all analysed the thematic associations between papers and patents from a content perspective. These articles were deep analyses into the internal mechanisms and interactions between science and technology at the micro level. Our framework also explores the relationship between science and technology from a content perspective. The correlations between the themes found are then used as a foothold to carry out collaborative innovation for a specific problem, as opposed to providing a general overview of a discipline.

Our review of the extant research thus far has focussed on scientific and technological relevance. However, beyond institutional competitiveness, institution type, and geo-relational indicators, research has also been undertaken from the perspective of the innovation chain. Although many studies have investigated the factors that influence partner selection with respect to innovation content, most only focus on a narrow aspect of innovation, such as basic research *or* technical development, not both. Alternatively, they focus on a single node of the innovation chain or the technological relatedness of inter-organisational links (Cloodt et al., 2006; Lane & Lubatkin, 1998; Xu et al., 2016). However, there is a risk that any collaboration based on such analysis may fall subject to intense competition. Hence, the selection framework we designed draws partners from different parts of innovation chain, as evidenced by an organisation's publication and patent portfolios.

Heterogeneity and complementarity

Bonaccorsi & Thoma (2007) find that both discovery and invention require structured interdependencies between institutions and that these are characterised by different goals (e.g., industry and academia). Additionally, Yang (2019) indicate that the essence of industry-university-research innovation is the complementarity of the knowledge structures of the three types of entities: enterprises, universities, and research institutes. Indeed, knowledge complementarity is also a key element in other forms of collaborative innovation, as is the source of knowledge synergies, being composed of the two dimensions of relevance and heterogeneity. Heterogeneity is the source of diversity; the greater the heterogeneity of knowledge between partners the more likely it is that those partners will successfully innovate. Knowledge relevance is the basis for communication and exchange, and this determines the probability that two partners will realise their innovation potential. When partners have similar basic knowledge and different expertise, synergy can exist, thus maximising the value of the partnership.

Dong et al. (2018) state that results of scientific research are reflected in scientific papers, which carry the basic scientific knowledge of humankind and permit solving the problems of 'what' and 'why'. Patent documents are the results of technological innovation and introduce numerous advanced technologies, which primarily permit solutions to 'how' problems. Further, from the perspective of the innovation process, scientific research institutes often conduct theoretical research to identify problems or propose theoretical solutions to them. For example, Hu et al. (2003) developed a rapid, specific, and sensitive TaqMan PCR method for detection of RSV (Respiratory Syncytial Virus) A and RSV B. Shi et al. (2008) developed the polymerase chain reaction (PCR) method for detecting unknown viruses, thus providing new ideas for the rapid detection of pathogens and emerging infectious diseases. Similarly, Liu et al. (2014) developed the chemiluminescence immunoassay (CLIA) method for detecting the hepatitis C virus (HCV) antibody, which has great clinical value for large-scale screening for HCV infections. Consequently, companies with production capacity or spin-off companies of scientific research institutes have since developed and produced reagent products based on these theoretical methods. For example, the paper of Hu et al. (2003) authored by Chinese University of Hong Kong provided support for (was cited by) one product patent (CN104593524, Nucleic acid detection kit for rapidly detecting respiratory syncytial viruses A and B and application of nucleic acid detection kit, granted on Jun 16, 2017) hold by Jiangsu Bioperfectus Technologies Co., Ltd. Therefore, this type of cooperation between research institutions and enterprises seamlessly connects knowledge flows from theory to production, which can accelerate the innovation process. Indeed, in a single paper, Liu et al. (2014) presents a body of cooperative research between a hospital and an enterprise. However, cooperative R&D between institutions not only produces papers and patent documents, but also results in products (or services) that can be used in practice to complete the process of innovation.

Hence, we assume that, for a specific problem, if a theoretical research institution publishes many related papers and a technology development and product manufacturing institution files many related patents, a strategic cooperation between the two institutions based on complementary advantages can achieve the aforementioned state of "being partners that simultaneously have similar basic knowledge but different expertise". Consequently, we have designed the framework to foster the process of transforming theoretical research into technology that solves problems (properly the first time), all the while avoiding intense competition to the extent possible.

That said, it is worth mentioning that the different types of research—scientific and technical—produce different types of results. In addition to these different purposes, papers are explicit; they usually describe the research process and results in detail, whereas the contents of patents are often simple and obscure to afford legal and economic protection. Therefore, finding a connection between different stages of the innovation chain based on papers and patent documents requires certain skills, such as deciphering different language styles. Nevertheless, although the two types of literature are very different, some clue words that characterise the common points of the research topics can still be found, with only this type of correlation signal needed to detect the possibility of cooperation between the two institutions. Thus, it is not necessary to thoroughly know the research content of the full text; however, the selection of subject words in topic interpretation is very important.

Our previous study (Qi et al., 2019) discusses the possibilities for cooperation between top-ranked organisations; however, cooperation can occur between any organisations. Moreover, the characteristics and cooperation possibilities of research

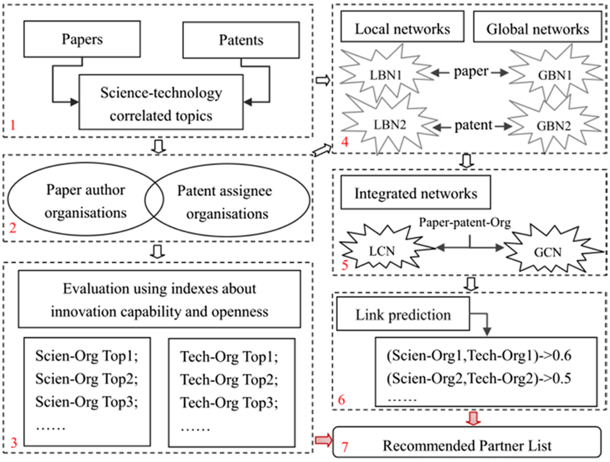

Fig. 1 Framework and process flow. Note: LBN1 and LBN2 indicate the basic local topic networks for papers and patents, respectively. Similarly, GBN1 and GBN2 indicate the basic global full-field networks for papers and patents. LCN and GCN are the local and global networks that integrate LBN1 and LBN2 → LCN and GBN1 and GBN2 → GCN

institutions with both scientific papers and technical patents in certain fields are worth analysing. For example, consider a list of topics in a field (e.g., A, B, C,…) and an institution (Org1) that holds patents in Topic A, but has not published any papers on this topic. However, it has published papers on Topic B, but has no related patents. Conversely, another institution (Org2) has published papers on Topic A but holds no related patents, while it has patents on Topic C but no related papers. If the two institutions cooperate on theoretical research and technology development on Topic A, it may be possible to improve the innovation efficiency of both.

Further, the innovation chain includes theoretical research, technology development, commercial production, and sales. Therefore, organisations that have both scientific papers and technical patents (sci-tech-org) in a field usually pay more attention to the integrity of innovation than do organisations that have only papers or patents, making collaborative innovation more efficient. It is well known that an effective method for predicting the cooperation of specific organisations is link prediction based on cooperative networks. Further, if this network is a fusion network of thesis and patent organisations, it can also screen for possible cooperation between organisations with published papers or patents. Therefore, this study draws on existing fusion methods and presents a collaborative innovation partner recommendation method that integrates topic analysis with link prediction.

Methodology

The framework of the method is shown in Fig. 1. The diagram shows two distinct processes. Steps 1–3 outline the text mining, topic analysis and indicator evaluation procedure (Part 1), while Steps 4–6 outline the network analysis and link prediction stages (Part 2). Step 7 brings the two parts together, offering the partnership recommendations.

Theoretical discussion

First, we investigated whether strategic cooperation on a particular innovation theme between heterogeneous institutions could help avoid intense competition by ensuring that each entity is located at a different point in the innovation chain. Heterogeneous institutions could mean one institution is good at scientific research—as evidenced by the fact that they have published many papers—while the other focuses on technological research and product manufacturing—as evidenced by the fact that they have numerous patents. Since cooperation is based on specific themes or problems, its purpose can be relatively explicit. If this is kept in mind, the success rate of cooperation and jointly solving problems can be relatively high. As such, the first step of the framework focuses on determining which science and technology topics are correlated. More specifically, scientific research topics are extracted from the papers and technical development topics are extracted from the patents. The scientific topics represent theoretical innovations, and the technical topics represent industrial applications. The next step is to draw correlations between the two lists, and this step is critical. Many criteria could be used: the co-occurrence of subject words could be examined, topic similarity operations could be employed, and so forth. Notably, many researchers involve field experts to manually interpret and evaluate the pathways from scientific invention to commercialisation.

Step 2 entails capturing the meta data of the documents, i.e., the authors, affiliations, and patentees, etc. Beginning with two separate sets of documents, divided by topics, Fig. 1 shows the two overlapping ovals. This is because some institutions may simultaneously have papers and patents on a topic. At this stage, every organisation associated with a given topic in either of the two groups is considered a candidate partner for collaborative innovation. According to the literature, some of the factors that influence good cooperation include strength matching, complementary resources, cultural compatibility, and whether there have been prior collaborations. These need to be assessed with indicators as is explained in the next step.

Many indicators already exist that can be reused or adapted. Among many others, these include strength, innovation ability, and attitude towards open innovation. One might also choose to develop bespoke indicators. In our framework, indicator evaluation is done on papers and patents independently. If an institution has both patents and papers, they will be evaluated separately, once with the paper data and once with the patent data. The next part of this process involves weighting and integrating the indicators into two separate organisation-topic lists. Here, each institution is given a score, termed the collaborative innovation index (CII) that reflects the ability and potential to take part in specific collaborative innovation. The authors and patentees are then sorted according to the CII, and each list is divided into three tiers: high, medium, and low. This is done on the principle that organisations in the same tier will find it easier to collaborate with each other. In practice, partners can be selected by referring to other factors if desired. Notably, if an institution appears in both lists at the same level, it indicates that the institution has equal capacity for theoretical research and technological development. These organisations should find they have greater power in the decision to collaborate with partners.

Step1: Identifying the topic correlations between papers and patents

The topics in a specific research field are extracted using a topic model, such as latent Dirichlet allocation (LDA), parallelLDA (PLDA), probabilistic latent semantic analysis

(PLSA) or any one of a number of alternative methods. Strongly associated topics are then determined based on a high level of similarity and expert judgements.

Extracting the topics

Topics can be extracted based on keywords or hierarchies. In our opinion, hierarchies offer a better reflection of innovation than keywords. However, the choice of which model to use may depend on many factors, such as text characteristics, the skills and know-how available. The most widely used topic model for semantic document mining is LDA. Typically, the corpus comprises the titles and abstracts of the documents and a national language processing (NLP) technique is used to extract precise and meaningful keywords from these fields. The input of the topic model is a list, which is taken to be a bag-of-words. The topic model then represents each document as an exchangeable bag-of-words. Since the quality of these bags-of-words is very important to the topic modelling result, it is highly advisable to use an inductive framework called 'term clumping' to clean them (Zhang et al., 2014). As the LDA model works, it generates a list of research topics that each comprise a discreet bag-of-words where each keyword is assigned a probability weight, i.e., the strength with which this keyword represents the topic. Further, each paper and patent document is represented as a bag-of-topics, again, each with their own probability weight (Blei et al., 2003). The PLDA model is a more recent extension of the LDA model, developed by Wang et al. (2009). Some believe it offers improved efficiency and precision over the traditional LDA.

Mining correlated topics

There are many kinds of correlations between science and technology, and, correspondingly, there are many correlated topics. One of the most feasible ways to find them is topic similarity. In theory, the greater the similarity between the topic content in a paper and a patent, the greater the likelihood they will be correlated. Hence, correlated topics can be found by applying topic similarity calculations to the corpus and supplemented with expert judgments. Following the LDA model, topics can be represented as:

$$\text{sim}(\text{topic}_i,\text{topic}_j) = \sum_{r=1}^{n}\sum_{k=1}^{m}\frac{p_{ir}\cdot p_{jk}\cdot \text{sim}(term_{ir}, term_{jk})}{m\cdot n} \tag{1}$$

where $p\left(term_k|topic_i\right)$ is the weight of the probability distribution of term_k in a given $\text{topic}_{i.}$

And the similarity $\text{sim}(\text{topic}_i, \text{topic}_j)$ between topic_i and topic_j can be calculated by:

$$\text{sim}\left(\text{topic}_i, \text{topic}_j\right) = \sum_{\text{r}=1}^{n}\sum_{k=1}^{m}\frac{p_{ir}+p_{jk}}{2}\cdot \text{sim}\left(term_{ir}, term_{jk}\right) \tag{2}$$

where n is the number of terms in topic_i; and m is the number of terms in $\text{topic}_{j.}$

Cosine similarity and other similarity analysis measures could also be used. Such measures are often based on a co-occurrence matrix of terms in the document set, and the topics pairs with similarities higher than a given threshold are regarded as correlated topics. Alternatively, topic pairs could be sorted by similarity and then examined and determined manually. Domain expertise is generally required in all cases because it is not necessarily certain that the higher the similarity, the more suitable the topic is for collaborative innovation.

Step 2: finding organisations

This step results in two lists of organisation names corresponding to the correlated topics: one for papers, the other for patents. Note that the institution names need to be disambiguated to ensure that different names for the same institution are consolidated into one standard spelling.

Step 3: selecting alternative organisations using an evaluation index

Criteria and indices

One correlated topic usually corresponds to multiple patents and papers, which in turn corresponds to multiple author institutions and patentees. Therefore, further evaluation and judgement is needed to determine which institutions pairs are most suitable for cooperation. Referring to existing studies (Geum et al., 2013; Zhang et al., 2015), we selected two criteria as our benchmarks: the institution's innovation capacity and their attitude towards open innovation.

Innovation capacity is usually measured by the number of papers and/or patents. We assume that filing patents and publishing papers represent achievements in different types of innovation, but, on the whole, the more innovative an institution is in either camp, the greater the innovation output in that type of innovation. Conversely, high quality output can reflect strong capacity. Therefore, we calculated five metrics based on papers and five metrics based on patents to reflect innovation capacity for each organisation and each topic as follows:

Innovation Capacity (IC)—Papers:

(1) Pu—The number of papers published by the attributed institution on this topic.
(2) PuI—The proportion of research papers published by the institution on this topic as compared to other topics. This proportion reflects the degree of importance that institution attaches to this topic. In theory, the more attention given to a topic, the stronger it is as a basis for cooperation. Similar to the index of "R&D emphasis" in Song et al. (2016).
(3) PuR—The share of papers published on a topic in comparison to other organisations. The larger the share, the stronger the candidate is and the greater its capacity.
(4) PuC—The influence of the institution on this topic, which is measured by the total number of citations received. This metric can also be limited to citations recevied within a specific period if desired.
(5) PuY—The number of years the institution has been publishing on this topic. In general, the greater the duration, the deeper the accumulation of knowledge and the easier it to achieve innovation.

Innovation Capacity (IC)—Patents:

(1) Pa—The number of patents the institution holds on a topic.
(2) PaI—The proportion of patents held on this topic as opposed to other topics.
(3) PaR—The share of patents held on this topic as compared to other organisations.
(4) PaC—The influence of the institution on this topic in terms of citations received.
(5) PaY—The number of years the institution has held patents on this topic.

Some studies have demonstrated that institutions that have cooperated in the past are more likely to cooperate again (e.g. Li & Rowley, 2002), whereas institutions that have not cooperated in the past may not be open to cooperation easily in future. For example, the larger the proportion of collaborative publications in an organisation's total set, the more willing that organisation is to collaborate. Further, the greater the number of partners the more open the institution's innovation policy. This gives rise to five more metrics based on papers and five metrics based on patents to reflect innovation openness for each organisation and each topic:

Innovation Openness (IO)—Papers:

(1) Puc—The number of co-authored papers on a topic (institution level).
(2) Pucr—The proportion of papers co-authored in this topic compared to other topics. This proportion reflects the degree of openness that institution attaches to this topic. In theory, the more open the attitude, the greater the probability of successful cooperation on the topic.
(3) PuCI—The number of partner institutions engaged on this topic. The greater the number of partners, the more open the institution's innovation policy on this topic.
(4) PuCY—The number of years the institution has been collaborating on this topic.
(5) PuCF—The average number of years the institution has been collaborating with different partners on this topic across all the partners.

Innovation Openness (IO)—Patents:

(1) Pac—The number of patents co-held on a topic.
(2) Pacr—The proportion of patents co-held in this topic compared to other topics.
(3) PaCI—The number of different partner institutions with whom co-patents are held.
(4) PaCY—The number of years the institution has been collaborating on this topic.
(5) PaCF—The average annual amount of cooperation.

Each dimension and its corresponding indicators are summarised in Table 1.

If the data corresponding to just the topic analysis is insufficient for some indexes, the model can be extended to consider the full literature of the field. This is because the purpose of the evaluation is solely to investigate the features of one organisation for collaborative innovation. In addition, the indicators devised for this study are only our recommendations; other indicators might be added in as needed to suit the practical application at hand. It is also worth noting that the results will reflect organisational characteristics across the entire span of the data period, so the period of data collected is also an important factor.

Table 1 Criteria and indices

Criteria	Index	Description
Scien-Org		
Innovation Capacity (IC)	Pu	The number of papers published by the institution on this topic
	PuI	The proportion of research papers published by the institution on this topic as compared to other topics
	PuR	The share of papers published on a topic in comparison to other organisations
	PuC	The influence of the institution on this topic, which is measured by the total number of citations received
	PuY	The number of years the institution has been publishing on this topic
Innovation Openness (IO)	Puc	The number of co-authored papers on a topic (institution level)
	Pucr	The proportion of papers co-authored in this topic compared to other topics
	PuCI	The number of partner institutions engaged on this topic
	PuCY	The number of years the institution has been collaborating on this topic
	PuCF	The average number of years the institution has been collaborating on this topic across all the partners
Tech-Org		
Innovation Capacity (IC)	Pa	The number of patents of the institution holds on a topic
	PaI	The proportion of patents held on this topic as opposed to other topics
	PaR	The share of patents held on this topic as compared to other organisations
	PaC	The influence of the institution on this topic in terms of citations received
	PaY	The number of years the institution has held patents on this topic
Innovation Openness (IO)	Pac	The number of patents co-held on a topic
	Pacr	The proportion of patents co-held in this topic compared to other topics
	PaCI	The number of different partner institutions with whom co-patents are held
	PaCY	The number of years the institution has been collaborating on this topic
	PaCF	The average annual amount of cooperation

Institute ranking and matching

We developed a single integrated index for conveniently sorting the candidate institutions, named CII, as noted above. As shown in Table 1, different weights are needed for different indices and criteria, and additionally the scores need to be normalised. The specific process and formula are based on Geum et al. (2013). The calculation models for CII are as follows:

$$\begin{aligned}&\text{Scien-Org}: \ \text{CII} = \text{W}_{11} * \left(\text{W}_{21} * \text{Puc}' + \text{W}_{22} * \text{Pucr}' + \text{W}_{23} * \text{PuCI}' + \text{W}_{24} * \text{PuCY}' + \text{W}_{25} * \text{PuCF}'\right)\\&+\text{W}_{12} * \left(\text{W}_{31} * \text{Pu}' + \text{W}_{32} * \text{PuI}' + \text{W}_{33} * \text{PuR}' + \text{W}_{34} * \text{PuC}' + \text{W}_{35} * \text{PuY}'\right),\end{aligned} \tag{3}$$

$$\begin{aligned}&\text{Tech-Org}: \ \text{CII} = \text{W}_{11}*\left(\text{W}_{21}*\text{Pac}' + \text{W}_{22}*\text{Pacr}' + \text{W}_{23}*\text{PaCI}' + \text{W}_{24}*\text{PaCY}' + \text{W}_{25}*\text{PaCF}'\right)\\&+\text{W}_{12}*\left(\text{W}_{31}*\text{Pa}' + \text{W}_{32}*\text{PaI}' + \text{W}_{33}*\text{PaR}' + \text{W}_{34}*\text{PaC}' + \text{W}_{35}*\text{PaY}'\right)\end{aligned} \tag{4}$$

where Puc' denotes the normalised value of Puc (the number of co-authored papers on a topic), and the rest follow by analogy. W_{11} and W_{12} are weights that consider the importance of the institution's innovation capability and openness to innovation. W_{2i} and W_{3j} are the weights that consider the importance of the number of publications, influence, year, and so forth. The classical analytic hierarchy process (AHP) method can be used to determine these weights. It should be noted that the weight values should be adjusted when selecting different indicators.

After obtaining the CII value, each institution is sorted and graded. Institutions can be matched by tier, geographical distance, or any other relevant factor depending on the nature of the study. It should be pointed out that these matches do not exclude those who have cooperated in the past or are already cooperating. Existing collaborations were not excluded for the simple reasons that: (a) they validate the method; and (b) they confirm the collaboration is worth proceeding with if there were any doubts. The final result is two ranked lists of organisations and topics—one for papers, the other for patents.

Step 4: constructing basic collaboration networks

The organisation collaboration networks for both papers (Paper-Org network) and patents (Patent-Org network) are constructed as basic local networks—LBN1 and LBN2, respectively. Each is derived from the literature of one topic. Two global cooperation networks (GBN1 and GBN2) are simultaneously built based on the full literature of the field so as to widely explore cooperation possibilities. Therefore, there are two arrows in the Fig. 1 framework diagram pointing to network construction (from Step 1 to 4 and from Step 2 to 4). Note that the full-field cooperation network (e.g., GBN1) will include the local network (e.g., LBN1) because there will be topic overlaps.

Step 5: converging the networks

Organisations that have both papers and patents on a particular topic are identified from the local networks LBN1 and LBN2 and consolidated into a list called Paper-Patent-Org in the framework. The same is done with the global networks GBN1 and GBN2. If an organisation does not appear on the local Paper-Patent-Org list then one might look to the global list through expert judgements to determine whether overarching field knowledge provides a good basis for cooperation on that particular topic. The purpose of finding these Paper-Patent-Org organisations is not only to identify the possible cooperation between them but also to construct the subsequent integrated network. Using link prediction in the next steps requires that we first build a converged comprehensive agency cooperation network composed of author institutions and patentees. Again, two networks are built—one local (LCN) and one global (GCN)—by merging the data matrices corresponding to the two basic networks. Two issues demand special attention during this integration. The first is how to integrate the two kinds of data—co-authors and co-patentees—and their relationships. The second is how to balance the different scales of the paper and patent networks. That is, it is far more common for organisations to cooperate on papers than it is on patents. So the average number of partners in the patent network LBN2 might only be 4 but that number might jump to 34 on the paper network LBN1. If the two matrices are pieced together directly, future connections between the science and technology institutions may not appear. Hence, data from the two basic networks—paper and patents—must be normalised to the same measurement standards before merging the data matrices. The construction process from two basic collaboration networks to an integrated collaboration network is shown in Fig. 2. The Paper-Patent-Org (the dual-coloured balls) are important intermediaries in the process, as these act as bridge organisations. As mentioned above, since they have both papers and patents and they cooperate with other institutions, they are simultaneously included in the paper network and the patent network, so they can be used as bridging nodes to combine the paper and patent networks into a whole network.

From the perspective of graph theory, the fusion process of the two networks can be simply described as follows. Suppose the paper cooperation network is $G1(V_1, E_1)$ where V_1 is the set of paper institution nodes, and the patent cooperation network is $G2(V_2, E_2)$ where V_2 is the set of patent institution nodes. There are the same institutions (nodes) in both networks. The merged network can be expressed as $G(V, E)$, where $V = V_1 \cup V_2$, edge set $E = E_1 \cup E_2$, and the weight of the edge is a linear combination of the cooperation frequency of the paper cooperation network and the patent cooperation network.

The data matrix can be simply described as follows. Suppose the cooperation matrix of thesis organisations is:

$$\mathrm{A} = \begin{bmatrix} X_{11} & \cdots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{m1} & \cdots & X_{mm} \end{bmatrix}$$

And the cooperation matrix of patent organisations is:

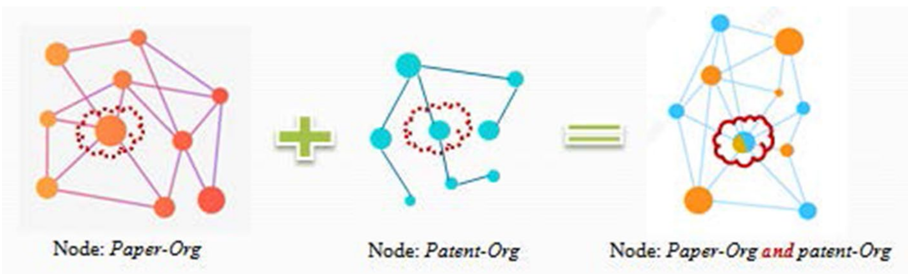

Fig. 2 The process of constructing an integrated network

$$\mathrm{B} = \begin{bmatrix} Y_{11} & \cdots & Y_{1n} \\ \vdots & \ddots & \vdots \\ Y_{n1} & \cdots & Y_{nn} \end{bmatrix}$$

The fusion matrix is:

$$\mathrm{G} = \left[W_1 * \mathrm{A},\ W_2 * \mathrm{B}\right] = \begin{bmatrix} Z_{11} & \cdots & Z_{1u} \\ \vdots & \ddots & \vdots \\ Z_{u1} & \cdots & Z_{uu} \end{bmatrix} \tag{5}$$

where $u = m + n\text{-}r$; r is the number of institutions that appear in the two matrices at the same time, and W is the weight used in data normalisation.

For the same organisation, the vectors from A and B will be merged as one vector:

$$\mathrm{Vector}\left(Z_{ij}\right) = \mathrm{Vector}\left(X_{ij'}\right) + \mathrm{Vector}\left(Y_{ij'}\right) \tag{6}$$

otherwise, the two vectors enter the fusion matrix respectively, as

$$\mathrm{Vector}\left(Z_{ij}\right) = \mathrm{Vector}\left(X_{ij'}\right) \text{ or } \mathrm{Vector}\left(Y_{ij'}\right) \tag{7}$$

Where $X_{ij'}$ and $Y_{ij'}$ denote the normalised value of X_{ij} and Y_{ij}. All vacant positions are filled with zeros.

Step 6: using the link prediction method

At last, we arrive at the link prediction step, which is performed on the integrated networks. It is with this step that the organisations with great potential for collaboration are identified. Link prediction is a huge topic with many different indicators that can be used to ascertain recommendations and many different models to do the actual predicting. Rather than discuss the strengths and weaknesses of each here, we refer readers to good quality existing link prediction research, such as Zhou et al. (2009), Yan & Gun (2014) and Yu et al. (2017).

Step 7: providing the recommended partner list

The final step is to review the predictions and generate a recommendation list in easy-to-read form. Here, all the analysis results of the previous steps should be comprehensively considered. The institutions and the possibilities for cooperation that were deduced in both Parts I and II of the procedure can be regarded as mutually confirming each other. However, we caution analysts not to disregard results that only appeared in one part and not the other because sometimes innovation requires change.

Case study

As our case study, we selected papers and patents in the field of Hepatitis C virus (HCV) research. In our previous research (Qi, 2019), we found that, in this field, there are many citations between patents and papers that display obvious continuity, indicating that the relationship between science and technology in this field is relatively strong. This makes HCV research a good field through which to verify our method. We retrieved 33,524 papers and 6804 patents from 2008 to 2017 from the Web of Science (WOS) and Derwent Innovations Index (DII) databases. Then, we divided them into five groups of two years each. Following our methodology, we used PLDA to extract topics from the papers and patents. For example, in the 2016–2017 group, 53 science topics were extracted from 6985 papers, and 104 technology topics were extracted from 975 patents.

Part 1: indicator ranking

By employing the above calculation process and manual judgement, we chose one of the related topics as an example to verify the feasibility of the method—HCV detection—which amounted to 430 papers and 347 author institutions, and 109 patents and 88 patentee institutions. These statistics give us publication volume (Pu) and patent volume (Pa). The number of partner institutions (PuCI) was adjusted to the average number of cooperative paper or patent institutions (set as PPuci). For example, if an institution had published 3 papers and cooperated with 6 institutions, the value of PPuci is 2. The CII value was calculated using Formula (9). Partial results of the institutional rankings are shown in Table 2.

$$\mathrm{CII} = 0.5*\left(\mathrm{Pu}'\,or\,\mathrm{Pa}'\right) + 0.5*\mathrm{PPuci}' \tag{8}$$

Table 2 Top 20 organisations

Rank	Scien-Org	CII	Tech-Org	CII
1	Univ Roma Tor Vergata	0.56	Hoffmann La Roche & Co Ag F	1.00
2	Cairo Univ	0.54	Caris Sci Inc	0.66
3	Hosp Univ Marques de Valdecilla	0.53	**Chinese Acad Sci**	0.66
4	Natl Taiwan Univ Hosp	0.53	Univ Korea Res & Business Found	0.66
5	Paris Descartes Univ	0.45	Univ Jinan	0.60
6	Inst Rech Dev IRD France	0.39	Roche Molecular Systems Inc	0.60
7	Univ Calgary	0.39	Guangzhou Tainuodi Biological Technology	0.60
8	AGS Norte Cadiz	0.37	Zhuhai TainuoMaibo Biotechnology Co Ltd	0.60
9	Erasmus MC	0.35	Servicio Andaluz Salud	0.58
10	**Mayo Clin**	0.35	Coyote Bioscience Co Ltd	0.56
11	Univ Seville	0.35	Birla Inst Technology & Sci Pilani	0.56
12	Fiocruz MS	0.33	Cent InvestigacionBiomedica Red	0.56
13	Curtin Univ	0.33	Gh Kansai Gakuin	0.56
14	San Francisco VA Med Ctr	0.33	Guy'S & St Thomas'S Hospital Nhs Trust	0.56
15	Mansoura Univ	0.32	India Dept Atomic Energy	0.56
16	Hosp Gen Univ Ciudad Real	0.31	Kings College London	0.56
17	Univ Paris Est	0.30	Mgm Inst Health Sci	0.56
18	**Johns Hopkins Univ**	0.29	Osaka Prefectural Hospital Org	0.56
19	Natl Sun Yat Sen Univ	0.29	Oxford Radcliffe Hospital Nhs Trust	0.56
20	Osaka Univ	0.28	**Univ Granada**	0.56

Institutions in bold reappear below Tables 3 and 4 as the result of link prediction procedure

From the table, we observe that most author institutions are colleges and universities, and that enterprises have the highest rankings in the 'Tech-Org' list. Strategic collaborations are likely to be fruitful between the top institutions in each list, i.e., the University of Rome Tor Vergata, Hoffmann La Roche & Co Ag F, Cairo University, and Caris Sci Inc. The remaining institution pairs might also cooperate, depending on any other factors considered in the screening.

In manually reviewing the portfolios of these organisations to validate our framework, we find that Hoffmann La Roche applied for some patents on techniques for detecting target nucleic acids, HCV genotype 3 in samples (e.g., blood), and a core polypeptide of HCV. They have also published some papers in collaboration with other institutions, including on the identification of HCV-resistant variants. One paper published by the University of Rome Tor Vergata evaluates the performance of a rapid method of quantifying HCV-RNA in the hepatic and extra hepatic compartments. Another paper evaluates the reliability and clinical utility of NS3 sequencing in HCV 1-infected patients. HCV-NS3 sequencing provides reliable results and, at the same time, it gives two clinically relevant pieces of information: a correct subtype/genotype assignment and the detection of variants that may interfere with the efficacy of protease inhibitor. Other papers evaluate the efficacy of all-oral direct-acting antiviral regimens, part of which requires implementing HCV-RNA detection or conducting NS3 and NS5A RASs tests by Sanger sequencing, and so on. The university does not hold patents on this topic; however, it could cooperate with Hoffmann La Roche to conduct theoretical R&D in this direction.

Cairo University holds no patents on HCV detection, but it has published some related research papers. For example, its publications include studies on phospholipidomic identification of potential serum biomarkers in dengue fever, hepatitis B and hepatitis C using liquid chromatography-electrospray ionisation-tandem mass spectrometry. In the studies, potentially dysregulated phosphatidylinositols (PLs) were considered as differentiating biomarkers to diagnose the diseases. In another study, researchers constructed a mathematical model for the early detection of hepatocellular carcinoma co-occurring with HCV infections. They also genotyped mir-221 and mir-101–1 related SNPS to investigate the association between mir-221 and mir-101–1 polymorphisms and their expressions. This can be used for the early prediction of hepatocellular carcinoma in HCV infected patients. Caris Sci Inc's patents are designed for many diseases. For example, they developed a new oligonucleotide used in composition for detecting microRNA-protein complex in a sample. They also developed a new oligonucleotide probe that is useful for characterising phenotypes and for diagnosing, prognosing, and treating cancer, autoimmune, cardiovascular, neurological, and infectious diseases. These are excellent grounds for Cairo University and Caris Sci Inc to consider collaborating.

Notably, some of the universities have spin-off companies. For instance, JiNan University has a spin-off called the Biopharmaceutical Research and Development Center of JiNan University, which is engaged in the R&D of genetically engineering drugs, biologically active peptides, and biological materials. JiNan University has applied for many patents on neutralising antibodies in collaboration with the Guangzhou Tainuodi Biotechnology Co Ltd and Zhuhai TainuoMaibo Biotechnology Co Ltd—for example, 'Preparation, detection and application of anti-HCV broad-spectrum neutralising antibody' (CN107286238). Therefore, cooperation between Cairo University and JiNan University could also involve enterprises to accelerate the innovation process. Of course, Cairo University could also cooperate with Hoffmann La Roche, and the University of

Rome Tor Vergata might cooperate with Caris Sci Inc or JiNan University. Additionally, three or more of these institutions can also work together to solve one problem, such as the development of improved diagnostics.

These suggestions or the possibility of cooperation are based on the content of the institution's past research and the abovementioned CII numericals. What is shows is that, even though the research content of a pair of institutions may not be exactly the same, in the future, joint R&D on certain research problems can be conducted based on actual clinical needs.

Part 2: link prediction

The first step in Part 2 is to construct the basic local and global networks. In comparing the basic local networks, we found that 6 institutions had both patents and papers: the University of Washington, the Chinese Academy of Sciences, the University of Grenoble Alpes, Abbvie Inc, the Institute of Salud Carlos III, and the University of Queensland. When constructing the basic global paper network, we found there was an excessive number of organisations, so we eliminated any organisation with a cooperation value of less than 2. (The maximum value was 46.) In the global patents network, there were only 728 patent agencies, with the maximum cooperation value of 6. In comparing the two basic global networks, we found 45 global Paper-Patent-Org institutions, including the aforementioned 6 institutions from the local list.

We analysed the papers and patents of these 6 institutions first. The University of Washington had one patent for a method of detecting HCV in a sample and had co-authored some papers on the detection of occult HCV infections with other institutes. The Chinese Academy of Science has published two papers on the detection and quantification of HCV RNA and one patent for aptamers used for detecting the HCV core antigen in patients. The papers and patents of the University of Grenoble Alpes focus on detecting HCV genotypes or accurate genotyping including recombinant form detection. Abbvie Inc has produced several papers on viral load tests and also holds some patents about detecting responsiveness to a direct-acting antiviral regimen and predicting the risks and effects of treatment failure. Two patents focus on detecting the methylation status of CpG islands in the promoter region of the IL28B gene and another focusses on assessing microRNA (miR)-122 levels. The University of Queensland's papers pay more attention to biomarkers, especially novel biomarkers to improve diagnostic accuracy. Its patents pertain to a biosensor used for detecting target molecules and diagnosing diseases or conditions in organisms. Institute of Salud Carlos III holds a patent detecting and quantifying the natural variants of the HCV genotype 1a and two papers co-authored with other Spanish institutions. One paper pertains to the detection of HCV core-specific antibodies, suggesting occult HCV infection among blood donors. The other is a new real-time-PCR method for identifying single point mutations in HCV.

Among the remaining 39 Paper-Patent-Org institutions, some institutions do not have any papers or patents on the topic of HCV detection; others only have papers. For example, the University of Strasbourg has published two papers titled: 'Signalome-wide assessment of host cell response to hepatitis C virus' and 'Tracking HCV protease population diversity during transmission'. The University of Miami has developed an on-site bundled rapid HIV/HCV test for use in prompt diagnosis of both infections and substance use disorders and treatment programmes. The University of Tokyo compared the predictive abilities of the Abbott real-time HCV assay (art) with those of standard serum HCV ribonucleic acid (RNA) detection methods. The research of University of Granada presents a method of cobas® HCV genotyping (gt) assay that correctly identifies HCV genotypes/subtypes 1a and 1b. Genentech Inc used the Genedrive® platform to develop an assay for the SNP rs12979860 variants (CC, CT and TT). It can be used for il-28b genotyping and may be useful for directing patients toward lower cost therapies. The University of British Columbia has evaluated the clinical performance of the core antigen (HCV-cAg) assay from plasma samples in monitoring the efficacy of HCV treatments and virus recurrence. They have also developed and validated a real-time, reverse transcription-PCR assay for rapid and low-cost genotyping of the HCV genotypes 1a, 1b, 2, and 3a, which has the potential to reduce cost and labour burdens in high-volume testing settings.

After standardising the data, we constructed the integrated networks and executed the link prediction procedure following some of the protocols outlined in Zhou et al. (2009). In their paper, Zhou and colleagues compare the performance of nine well-known local node similarity measures on six real networks and, based on their findings, they propose two new indicators: the RA Index (Resource Allocation) and the LP Index (Local Path). Both have significantly better predictive capabilities than the nine existing indicators. The RA index considers a pair of nodes (x and y) that are not directly connected but where the similarity between x and y is defined as the amount of resource y receives from x:

Table 3 Partial link prediction results

Source	Org1	Org2	Probability of cooperation
GCN	Kagawa Prefectural Cent Hosp	*Univ Tokyo*	0.71
	Univ Washington	Massachusetts Gen Hosp	0.65
	Univ Miami	Johns Hopkins Univ	0.56
	Hosp Univ Miguel Servet	Inst Salud Carlos III	0.53
	Univ Tokyo	Kagoshima Univ	0.53
	Ogaki Municipal Hosp	Univ Tokyo	0.53
	Univ Barcelona	Univ Pompeu Fabra	0.50
	Univ Grenoble Fourier Joseph	Univ Grenoble Alpes	0.50
	Seoul Natl Univ	Univ Chonnam Nat Ind Found	0.50
	Chinese Acad Sci	Univ Texas Md Anderson Canc Ctr	0.50
	Univ Washington	Va Greater Los Angeles Healthcare Syst	0.50
	Univ Maryland	Massachusetts Gen Hosp	0.49
	Univ Maryland	Harvard Med Sch	0.48
	Mayo Clin	Univ Washington	0.48
	Aarhus Univ Hosp	Univ Plymouth	0.47
	Univ Strasbourg	Hannover Med Sch	0.47
	Univ Calif San Francisco	Univ Michigan	0.45
	Icahn Sch Med Mt Sinai	Univ Maryland	0.43
	Univ Sydney	Royal Melbourne Hosp	0.43
	Rockefeller Univ	Inst Salud Carlos III	0.43
LCN	Kyoto Univ	Abbvie Inc	0.33
	Univ Queensland	Royal Melbourne Hosp	0.23

$$S_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k(z)} \tag{9}$$

For a given node x, $\Gamma(x)$ denotes the set of neighbours of x, and $k(z)$ denotes the degree of node z.

Some of the pairs with higher cooperation probabilities are shown in Table 3. The institutions underlined against a green background are the six featured institutions with both papers and patents on HCV detection, and the institutions in red italics have papers on this topic.

There is less data in the local networks, so the links in the local integrated network (LCN) are relatively sparser. Also, the probability values are smaller but more specific to each topic. There is more data in the global networks, and so there are more results for GCN. However, some of the recommended collaborations relate to topics other than detection, such as drug development. A quick solution to cut through the large set of results is to find the institutions that have papers and/or patents on the correlated topic of HCV detection, such as our six featured institutions. The global integrated network shows us that four of these six appear in the first 20 prediction results (20/188), while Table 4 shows us some other notable results. The institutions underlined against a green background are the two of six featured institutions with both papers and patents on HCV detection, and the institutions in red italics have papers on this topic.

It needs to be pointed out that the prediction results of the two integrated networks (GCN vs LCN) should be treated differently because the data values of the two integrated networks are not normalised for ranking. Therefore, the prediction results of two integrated networks in Table 3 cannot be sorted simply according to the probability of cooperation. Further, the prediction result of the LCN cannot be ruled out simply because the values are small. At the same time, we found that the papers of some institutions may not propose new detection methods or markers. Rather, they merely apply or evaluate existing methods. Cooperating with these institutions may result in a fast track to innovative detection methods, and they can provide relevant materials, such as evidence as to the clinical effect of testing reagents or equipment and screening strategies for specific populations. The cooperation we envision is that two or more institutions will conduct theoretical and design research on detection methods and the development of detection reagents or equipment based on clinical needs. Further, the results of clinical usage can be fed back to improve methods and technologies.

With this in mind, we manually analysed the prediction results to assess the effectiveness of the method. Consider some of the concrete examples in Table 3. Pair A—University of Tokyo and Kagawa Prefectural Central Hospital—co-authored a paper demonstrating that tests for the TLL1snp (single nucleotide polymorphisms) might be used to identify patients at risk of hepatocellular carcinoma. We believe that they will continue to cooperate

Table 4 Part of prediction results of the remaining several institutions in the global network

Org1	Org2	Probability of cooperation
Univ Florida	*Univ British Columbia*	0.40
Univ Paris Diderot	*Genentech Inc*	0.37
Inst Investigacions Biomediques Sunyer	*Univ Granada*	0.37
Abbvie Inc	Univ Calif San Francisco	0.36
Burnet Inst	Univ Queensland	0.33

to test their theories. At the same time, we suggest that University of Tokyo apply for new patents in this area, or they could cooperate with other organisations that already have patents in this area, such as Institute of Salud Carlos III or Abbvie Inc, in a tripartite or multiparty collaboration.

Pair B—The University of Washington and Massachusetts General Hospital—co-published a paper titled 'Hepatitis-C-virus-induced microRNAs dampen interferon-mediated antiviral signaling'. Further, a series of papers by Massachusetts General Hospital describe the increasing demand for HCV testing in specific populations; they quantify the genotype 3 HCV RNA; and so on. Given the abovementioned Univ Washington's papers and patents, these two organisations are therefore a likely match for collaborating on HCV detection.

Pair C—The University of Miami and Johns Hopkins University—have not collaborated. Miami has published one paper on on-site bundled rapid HIV/HCV testing, and Johns Hopkins has some research on HIV/HCV co-infected people. Moreover, Johns Hopkins has also published research on biomarkers, e.g., using HCV immunoglobulin G antibody avidity as a biomarker to estimate the population-level incidence of HCV infection. Therefore, these two institutions could collaborate on HCV or HIV/HCV testing and may consider developing and producing new detection reagents in partnership with private enterprises.

Pair D—Institute of Salud Carlos III and Miguel Servet University Hospital—coauthored a paper on the prevalence and patient characteristics of HIV/HCV co-infection in Spain. Other papers by Miguel Servet concern the prevalence and distribution of HCV genotypes in Spain. Salud Carlos III has both papers and patents and could work with Miguel Servet to perform research and to test, apply, evaluate, and improve the detection methods developed by the two organisations.

Pair E—Kyoto University and Abbvie Inc—both have interests in microRNA as it pertains to HCV. The papers published by Kyoto University focus on novel entry pathways for HCV (e.g., very-low-density lipoprotein receptor to mediate HCV entry independent of cd81) and the differential expression of microRNA in the liver of HIV/HCV co-infections. Abbvie Inc's patents involve the detection of microRNA (miR)-122 levels. Thus, it would be reasonable for these two institutions to engage in in-depth research collaborations on this topic.

By comparing the results of the link prediction process with the CII evaluation, we find that some organisations, such as the Chinese Academy of Sciences, Johns Hopkins University, the Mayo Clinic, and the University of Granada appear in both Table 2 and Tables 3 or 4. Their CII values are not very high, and this confirms our hypothesis that cooperation does not always happen between the top-ranked organisations. Moreover, there are fewer patentees in the link prediction results than in the CII evaluation results due to the relatively sparse connections between the author institutions and the patentees in the integrated networks. Recall that, here, that the Paper-Patent-Org list comprises important intermediaries. Therefore, both results of the link prediction and CII evaluation are included in the final recommended partner list. And, the results of both parts can be considered comprehensively. For example, the author institutions from the link prediction results, such as the University of Miami, can cooperate with the Hoffmann La Roche & Co Ag F, Caris Science Inc, the University of Jinan, and other institutions listed in the CII results.

Discussion and conclusions

Based on the literature and our previous analysis of the association between science and technology, we put forward a method for selecting collaborative innovation partners. The framework leverages text mining, topic analysis, network analysis, and link prediction with both research articles and patents to identify potentially fruitful science-technology relationships all along the innovation chain. The framework works best when targeted at specific scientific problems as opposed to providing an overall view of a field. It also excels at recommending organisations for different nodes of the innovation chain. Theoretically, it should improve the success rate of collaborations, helping to avoid intense competition because institutions with different innovation skills can form an innovation ecosystem with complementary knowledge and harmonious symbiosis. Innovation organisations can communicate and create new knowledge based on a common knowledge base and complementary skills, while forming stable, long-term relationships. This would accelerate the innovation process and application rate of scientific achievements, while enhancing the partner's innovation ability and competitiveness.

In the present case study, we used content analysis to investigate and demonstrate the effectiveness of the CII evaluation and link prediction results, and we compared them both. The common points of the two parts mutually verify each other, and the differences complement each other as a reference. Through a manual analysis of the literature, we found that the framework's recommendations proved to be either fortuitous and highly promising, or a reflection of collaborations that have already come to pass. Both provide confirmation that this method can serve as a valuable reference for policymakers regarding innovation cooperation among research institutions and/or high-tech enterprises. It also reflects the feasibility, validity, and robustness of the method. Although most institutions are usually aware of the R&D situation in other institutions within a similar field, such methods can still provide quantitative results from the overall perspective of the field as evidence. In fact, our proposed approach is more of a systematic framework because the choice of specific evaluation indices can be adjusted to suit the local conditions in practice. This suggested framework is expected to be valuable as a complementary tool for decision-making on R&D collaboration.

In summary, we believe that by combining the internal and external characteristics of papers and patents derived from the association between science and technology, and the innovation chain theory, our new method has expanded the theoretical understanding and practical application of collaborative innovation, R&D cooperation, and link prediction. It can also provide a reference for follow-up research on these topics.

Key notes

In this method, the most important step is to correctly determine which topics are related. When the result of text mining and semantic computing by NLP techniques is not ideal, properly interpreting and selecting topics may require domain experts. This is especially so when attempting to identify and process the correlated topics worth cooperating on. In our case study, although we read of many different detection methods, test objects, and specific research directions, we did not refine the correspondence between research institutions for each detection method and research point. Rather, we wanted to consider wide possibilities for cooperation. After the subject of study is determined, the scope of the paper and patent collection corresponding to the topics is also critical. Thus, it may be necessary to conduct extended collection, instead of being limited to a list of articles given by a program. Just like any information retrieval, the documents collected on a topic will generally cover slightly different aspects of the phenomenon and the content may or may not be accurate. Therefore, a manual interpretation of the results is indispensable, especially the analysis of research content. In our case study, the research content of some institutions may only contribute to the scientific field, and cooperation with them may not be as efficient or effective as with the institutions paying more attention to related technological problems. Finally, other aspects of the method are also critical, such as retrieving a complete corpus of papers and patents and ensuring that the literature is relevant to the field under study. Other important factors to consider include the time range, the processing effect of different LDA tools, data clumping, criteria and indices selection, the weights used in the CII formula, and the strategy used to standardise data with large differences in value ranges from the two basic cooperation networks into the integrated networks. All of these aspects will affect the correctness and the objectivity of the final result.

Limitations and future work

Our study has the following limitations. Based on the empirical process, our fusion method only has a certain level of feasibility. The link prediction method can provide specific cooperation probability values to help quantify judgements, while previous evaluation indices can identify more technical R&D institutions and other cooperation possibilities. Thus, the fusion method proposed in this article is suitable for comprehensive recommendation. However, since we only used two years' worth of empirical data, our analysis results might not be sufficiently representative. Further application and verification of the method must be conducted using larger amounts of data over longer periods of time. In the future, we must also consider each step more comprehensively, especially the determination of correlated topics, as there are many other metrics and methods. Finally, follow-up studies should examine the rationality of the proposed indicators, along with the correlations between those indicators, to improve the accuracy of the indicator system and the link prediction.

Acknowledgements The work was supported by the Fundamental Research Funds for the Central Universities—PUMC Youth Fund (Grant No. 2017330008) and the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (Grant No. 2017PT63008). In addition, we thank the anonymous reviewers for valuable comments and suggestions which helped us to improve this paper.

References

Alexiev, A. S., Volberda, H. W., & Van den Bosch, F. A. J. (2016). Interorganizational collaboration and firm innovativeness: Unpacking the role of the organizational environment. *Journal of Business Research, 69*(2), 974–984.

Ardito, L., Albino, V., & Messeni Petruzzelli, A. (2019a). The innovative implications of alliance ambidexterity and the moderating role of firm age. *IEEE Transactions on Engineering Management*, in press.

Ardito, L., Galati, F., Messeni Petruzzelli, A., & Corvino, A. (2019b). Improving the financial performance of SMEs. *Business Process Management Journal, 26*(5), 1041–1054.

Ariño, A., Abramov, M., Skorobogatykh, I., Rykounina, I., & Vila, J. (1997). Partner selection and trust building in west European-Russian joint ventures. *International Studies of Management and Organization, 27*(1), 19–37.

Autant, B. C. (2001). Science and knowledge flows: Evidence from the French case. *Research Policy, 30*(7), 1069–1078.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99–120.

Bassecoulard, E., & Zitt, M. (2004). Patents and publications: The lexical connection. *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems* (pp. 665–694). Kluwer Academic Publishers.

Becker, W., & Dietz, J. (2004). R&D cooperation and innovation activities of firms—Evidence for the German manufacturing industry. *Research Policy, 33*(2), 209–223.

Bhattacharya, S., Kretschmer, H., & Meyer, M. (2003). Characterizing intellectual spaces between science and technology. *Scientometrics, 58*(2), 369–390.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Bonaccorsi, A., & Thoma, G. (2007). Institutional complementarity and inventive performance in nanoscience and technology. *Research Policy, 36*(6), 813–831.

Brouthers, K. D., Brouthers, L. E., & Wilkinson, T. J. (1995). Strategic alliances: Choose your partners. *Long Range Planning, 28*(3), 18–25.

Cao, X., & Song, Q. (2016). The selection of I-U cooperative partner based on industry's QFD and improved VIKOR. *Science and Technology Management Research, 8*, 91–97. In Chinese.

Capaldo, A., & Petruzzelli, A. M. (2014). Partner geographic and organizational proximity and the innovative performance of knowledge-creating alliances. *European Management Review, 11*, 63–84.

Carayannis, E. G., Kassicieh, S. K., & Radosevich, R. (2000). Strategic alliances as a source of early-stage seed capital in new technology-based firms. *Technovation, 20*(11), 603–615.

Carpenter, M. P., & Narin, F. (1983). Validation study: Patent citations as indicators of science and foreign dependence. *World Patent Information, 5*(3), 180–185.

Chen, J., & Yang, Y. J. (2012). Theoretical basis and content for collaborative innovation. *Studies in Science of Science, 32*(2), 161–164. In Chinese.

Chen, W., Qu, H., & Chi, K. (2021). Partner selection in China interorganizational patent cooperation network based on link prediction approaches. *Sustainability, 13*(2), 1003.

Chesbrough, H. (2003). Open innovation: How companies actually do it? *Harvard Business Review, 81*(7), 12–14.

Clauss, T., & Kesting, T. (2017). How businesses should govern knowledge-intensive collaborations with universities: An empirical investigation of university professors. *Industrial Marketing Management, 62*, 185–198.

Cloodt, M., Hagedoorn, J., & Van Kranenburg, H. (2006). Mergers and acquisitions: Their effect on the innovative performance of companies in high-tech industries. *Research Policy, 35*(5), 642–654.

Cui, J., Ding, W. Q., Ma, J. H., & Yuan, H. M. (2021). Identification and selection of R&D partners based on technology concentric diversification: Taking Chinese medicine enterprises as an example. *Science and Technology Management Research, 4*, 133–139. In Chinese.

Dacin, M. T., Hitt, M. A., & Levitas, E. (1997). Selecting partners for successful international alliances: Examination of US and Korean firms. *Journal of World Business, 32*(1), 3–16.

De Solla Price, D. J. (1965). Is technology historically independent of science? A study in statistical historiography. *Technology and Culture, 6*(4), 553–568.

Ding, J. D., & Guo, J. (2021). Mining potential author cooperative relationships based on the similarity of content and path. *Information Studies: Theory & Application, 44*(1), 124-128+123. In Chinese.

Dong, K., Xu, H. Y., Luo, R., Wang, C., & Fang, S. (2018). Review of the research on relationship between science and technology. *Journal of the China Society for Scientific and Technical Information, 37*(6), 642–652.

Doz, Y. (1996). The evolution of cooperation in strategic alliances: Initial conditions or learning processes? *Strategic Management Journal, 17*(S1), 55–83.

Elia, S., Messeni Petruzzelli, A., & Piscitello, L. (2019). The impact of cultural diversity on innovation performance of MNC subsidiaries in strategic alliances. *Journal of Business Research, 98*, 204–213.

Etzkowitz, H., Webster, A., Gebhardt, C., & Terra, B. R. C. (2000). The future of the university and the university of the future: Evolution of ivory tower to entrepreneurial paradigm. *Research Policy, 29*(2), 313–330.

Friesike, S., Widenmayer, B., Gassmann, O., & Schildhauer, T. (2015). Opening science: Towards an agenda of open science in academia and industry. *The Journal of Technology Transfer, 40*(4), 581–601.

Garfield, E. (1984). Patent citation indexing and the notions of novelty, similarity and relevance. *Essays of an Information Scientist, 7*(3), 536–542.

Garg, K. C. (2001). A study of collaboration in laser science and technology. *Scientometrics, 51*(2), 415–427.
Geum, Y., Lee, S., Yoon, B., & Park, Y. (2013). Identifying and evaluating strategic partners for collaborative R&D: Index-based approach using patents and publications. *Technovation, 33*(6–7), 211–224.
Gnyawali, D. R., & Park, B.-J. (2011). Co-opetition between giants: Collaboration with competitors for technological innovation. *Research Policy, 40*(5), 650–663.
Guns, R., & Rousseau, R. (2014). Recommending research collaborations using link prediction and random forest classifiers. *Scientometrics, 101*(2), 1461–1473.
Hagedoorn, J. (1993). Understanding the rationale of strategic technology partnering: Interorganizational modes of cooperation and sectoral differences. *Strategic Management Journal, 14*(5), 371–385.
Hagedoorn, J. (2002). Inter-firm R&D partnerships: An overview of major trends and patterns since1960. *Research Policy, 31*(4), 477–492.
Hagedoorn, J., Roijakkers, N., & Kranenburg, H. V. (2008). The formation of subsequent inter-firm R&D partnerships between large pharmaceutical companies and small, entrepreneurial biotechnology firms – how important is inter-organisational trust? *International Journal of Technology Management, 44*(1–2), 81–92.
Han, J., Tang, X., & Yu, A. L. (2021). Research on identification of potential partnership based on link prediction with multilayer networks. *Systems Engineering-Theory & Practice, 41*(4), 1049–1060. In Chinese.
Hitt, M. A., Dacin, M. T., Levitas, E., Arregle, J. L., & Borza, A. (2000). Partner selection in emerging and developed market contexts: Resource-based and organizational learning perspectives. *Academy of Management Journal, 43*(3), 449–467.
Hu, A., Colella, M., Tam, J. S., Rappaport, R., & Cheng, S.-M. (2003). Simultaneous detection subgrouping and quantitation of respiratory syncytial virus A and B by real-time PCR. *Journal of Clinical Microbiology, 41*(1), 149–154. https://doi.org/10.1128/JCM.41.1.149-154.2003.
Huang, S. J. (2010). *Report on the innovation and development of China's industrial clusters: To build the innovative capabilities*. Economy & Management Publishing House.
Inkpen, A. C., & Currall, S. C. (1997). International joint venture trust: An empirical examination. In B. W. Beamish & J. P. Killing (Eds.), *Cooperative strategies: North American perspectives* (pp. 308–334). New Lexington Press.
Ionescu, A.-F., & Vernic, R. (2021). MOSOSS: An adapted multi-objective symbiotic organisms search for scheduling. *Soft Computing, 25*(14), 9591–9607.
Ireland, R. D., Hitt, M. A., & Vaidyanath, D. (2002). Alliance management as a source of competitive advantage. *Journal of Management, 28*(3), 413–446.
Jeon, J., Lee, C., & Park, Y. (2011). How to use patent information to search potential technology partners in open innovation. *Journal of Intellectual Property Rights, 16*(5), 385–393.
Ji, G. J., Yu, M. H., & Tan, K. H. (2020). Cooperative innovation behavior based on big data. *Mathematical Problems in Engineering*, Article ID 4385810.
Ju, N. Q., & Liu, G. J. (2004). On the essence and requirement in transforming scientific advances into technological productive force from the relations between science and technology. *Journal of Henan University (social Science), 44*(1), 24–28. In Chinese.
Jugend, D., Fiorini, P. D. C., Armellini, F., & Ferrari, A. G. (2020). Public support for innovation: A systematic review of the literature and implications for open innovation. *Technological Forecasting and Social Change, 156*, 119985.
Kale, P., Singh, H., & Perlmutter, H. (2000). Learning and protection of proprietary assets in strategic alliances: Building relational capital. *Strategic Management Journal, 21*(3), 217–237.
Kang, J., Lee, J., Jang, D., & Park, S. (2019). A methodology of partner selection for sustainable Industry-University cooperation based on LDA topic model. *Sustainability, 11*(12), 3478.
Lane, P. J., & Lubatkin, M. (1998). Relative absorptive capacity and interorganizational learning. *Strategic Management Journal, 19*(5), 461–477.
Laursen, K., & Salter, A. (2006). Open for innovation: the role of openness in explaining innovation performance among U.K. manufacturing firms. *Strategic Management Journal, 27*(2), 131–150.
Leiponen, A., & Helfat, C. E. (2010). Innovation objectives, knowledge sources, and the benefits of breadth. *Strategic Management Journal, 31*(2), 224–236.
Li, M. F., Liu, G. X., & Liu, L. (2011). Empirical comparative study on the participation of enterprises and universities in Industry-university-research institute collaboration: Collaborative content, level and mode. *R&D Management, 23*(4), 113–118. In Chinese.
Li, S. X., & Rowley, T. (2002). Inertia and evaluation mechanisms in interorganizational partner selection: Syndicate formation among US investment banks. *Academy of Management Journal, 45*(6), 1104–1119.

Liben-Nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology, 58*(7), 1019–1031.

Littler, D., Leverick, F., & Bruce, M. (1995). Factors affecting the process of collaborative product development: A study of UK manufacturers of information and communications technology products. *Journal of Product Innovation Management, 12*(1), 16–32.

Liu, H. B., & Wang, C. M. (2005). Enterprises strategy alliance and model for cooperative partner appraising and choosing. *Science Technology and Industry, 5*(12), 21-23+27. In Chinese.

Liu, J., & Sun, W. (2017). Discovery of potential scientific and technical collaborative relationship based on link prediction. *Information Studies: Theory & Application, 40*(7), 88-921+21. In Chinese.

Liu, L. J., Xie, Y. L., Dai, Z. X., Zhuo, C. S., & Wu, Y. S. (2014). Establishment and clinical application of chemiluminescence immunoassay for Hepatitis C Virus antibody. *Letters in Biotechnology, 25*(1), 97–99. In Chinese.

Liu, Z. Q., Xu, H. Y., Luo, R., Dong, K., & Zhu, L. J. (2019). Research on scientific and technological interaction patterns based on topic relevance analysis. *Journal of the China Society for Scientific and Technical Information, 38*(10), 997–1011. In Chinese.

Lo, S. S. (2009). Scientific linkage of science research and technology development: A case of genetic engineering research. *Scientometrics, 82*(1), 109–120.

Makri, M., Hitt, M. A., & Lane, P. J. (2010). Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions. *Strategic Management Journal, 31*(6), 602–628.

Maraut, S., & Martínez, C. (2014). Identifying author–inventors from Spain: Methods and a first insight into results. *Scientometrics, 101*(1), 445–476.

Meyer, M. (2000). Does science push technology? Patents citing scientific literature. *Research Policy, 29*(3), 409–434.

Narin, F., Hamilton, K. S., & Olivastro, D. (1997). The increasing linkage between US technology and science. *Research Policy, 26*(3), 317–330.

Nguyen, H.-K. (2021). Applications optimal math model to solve difficult problems for businesses producing and processing agricultural products in vietnam. *Axioms, 10*(2), 90.

Nielsen, B. B. (2003). An empirical investigation of the drivers of international strategic alliance formation. *European Management Journal, 21*(3), 301–322.

Ottonicar, S. L. C., Arraiza, P. M., & Fabiano, A. (2020). Opening science and innovation: Opportunities for emerging economies. *Forsajt (форсайт), 14*(4), 95–111.

Park, I., Jeong, Y., Yoon, B., & Mortara, L. (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.

Petruzzelli, A. M. (2011). The impact of technological relatedness, priorties, and geographical distance on university-industry collaborations: A joint-patent analysis. *Technovation, 31*(7), 309–319.

Qi, Y. (2019). Research on the analytical method of science technology relationship evolution in specific fields based on patent citation analysis——Taking HCV as an example. *Library and Information Service, 63*(23), 97–105. In Chinese.

Qi, Y., Hu, Z. Y., Xiang, B., Liu, C. J., Xu, H. Y., & Wen, Y. (2019). Identifying and evaluating strategic partners for collaborative innovation: A method based on a topic analysis of papers and patents, In *Proceedings of the international conference on scientometrics and informetrics (ISSI2019)* (pp. 2744–2745), II.

Rajalo, S., & Vadi, M. (2021). Collaboration potential between low-capacity SMEs and academic researchers determined by symmetry of motivation. *Technovation, 107*, 102304.

Rothaermel, F. T. (2001). Incumbent's advantage through exploiting complementary assets via inter firm cooperation. *Strategic Management Journal, 22*(6–7), 687–699.

Schmoch, U. (1997). Indicators and the relations between science and technology. *Scientometrics, 38*(1), 103–116.

Shi, B. S., Zhang, X. N., Gu, S. M., Tian, D., Yuan, Z. H., & Hu, Y. W. (2008). Establishment of a random PCR method for detecting unknown virus. *Journal of Microbes and Infections, 3*, 134-137+146. In Chinese.

Song, B., Seol, H., & Park, Y. (2016). A patent portfolio-based approach for assessing potential R&D partners: An application of the Shapley value. *Technological Forecasting & Social Change, 103*, 156–165.

Sun, X. L., & Ding, K. (2018). Identifying and tracking scientific and technological knowledge memes from citation networks of publications and patents. *Scientometrics, 116*(3), 1735–1748.

Tian, B. L. (2014). *Research on strategic alliance partner selection of high-tech enterprises*. Yangzhou University. In Chinese.

Tran, D.-H., Cheng, M.-Y., & Prayogo, D. (2016). A novel multiple objective symbiotic organisms search (MOSOS) for time-cost-labor utilization tradeoff problem. *Knowledge-Based Systems, 94*, 132–145.

University of California. (2015, September 15). What is continuous innovation. Retrieved January 7, 2022, from https://blink.ucsd.edu/sponsor/innovate/whatis.html#:~:text=Noun%3A%20modest%2C%20incremental%2C%20ongoing,activities%2C%20operations%2C%20and%20creations.

Verbeek, A., Debackere, K., & Luwel, M. (2003). Science cited in patents: A geographic 'flow' analysis of bibliographic citation patterns in patents. *Scientometrics, 58*(2), 241–263.

Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E., & Deleus, F. (2002). Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics, 54*(3), 399–420.

Veugelers, R. (1997). Internal R&D expenditures and external technology sourcing. *Research Policy, 26*(3), 303–315.

Wang, G. B., & Guan, J. C. (2011). Measuring science-technology interactions using patent citations and author–inventor links: An exploration analysis from Chinese nanotechnology. *Journal of Nanoparticle Research, 13*(12), 6245–6262.

Wang, H. H., Xie, F. J., & Zhou, S. (2014). 'Four-dimensional' collaborative framework for implementing innovation-driven development strategy in China from the perspective of innovation ecosystem. *Science & Technology Progress and Policy, 31*(17), 7–11. In Chinese.

Wang, J. (2016). *Research on partner selection of industry-university-research-application in the Internet of things*. Shanghai University of Engineering Technology. In Chinese.

Wang, J. F., Yang, S. S., & Wang, Y. D. (2016). Research on ecological innovation development path of technology park based on grounded theory. *Science & Technology Progress and Policy, 33*(18), 77–85. In Chinese.

Wang, L., & Huang, Z. (2020). Research on the synergetic innovation between pharmaceutical enterprises and scientific research institutions based on the quantum game. *IEEE Access, 8*, 63718–63724.

Wang, X. P. (2020). *Research on the identification and evolution of science-technology topic association*. Wuhan University. In Chinese.

Wang, X. F., Fu, Y., Qiu, P. J., & Liu, Y. Q. (2015). Identification of R&D partners based on SAO analysis. *Science Research Management, 36*(10), 19–27. In Chinese.

Wang, X. F., Qiao, Y. L., Hou, Y. J., Zhang, S., & Han, X. T. (2021). Measuring Technology Complementarity Between Enterprises With an hLDA Topic Model. *IEEE Transactions on Engineering Management, 68*(5), 1309–1320.

Wang, X. F., Wang, Z. N., Huang, Y., Liu, Y. Q., Zhang, J., Heng, X. F., & Zhu, D. H. (2017). Identifying R&D partners through Subject-Action-Object semantic analysis in a problem & solution pattern. *Technology Analysis & Strategic Management, 29*(10), 1167–1180.

Wang, Y., Bai, H., Stanton, M., Chen, W. Y., & Chang, E. Y. (2009). PLDA: Parallel latent dirichlet allocation for large-scale applications. *Lecture Notes in Computer Science, 5564*, 301–314.

Wang, Z. B., Han, W. M., Sun, Z. M., & Pan, X. L. (2019). Research on scientific collaboration prediction based on the combination of network topology and node attributes. *Information Studies: Theory & Application, 42*(8), 116-120+109. In Chinese.

Wen, J., & Kobayashi, S. (2001). Exploring collaborative R&D network: Some new evidence in Japan. *Research Policy, 30*(8), 1309–1319.

Wu, W. Y., Shih, H. A., & Chan, H. C. (2009). The analytic network process for partner selection criteria in strategic alliances. *Expert Systems with Applications, 36*(3), 4646–4653.

Xu, H. Y., Wei, L., Pang, H. S., Guo, T., & Fang, S. (2016). Methods to identify potential Industry-University-Research institutions cooperation partners. *Journal of the China Society for Scientific and Technical Information, 35*(5), 521–529. In Chinese.

Xu, S., Zhai, D. S., Wang, F. F., An, X., Pang, H. S., & Sun, Y. R. (2019). A novel method for topic linkages between scientific publications and patents. *Journal of the Association for Information Science & Technology, 70*(9), 1026–1042.

Yan, E., & Guns, R. (2014). Predicting and recommending collaborations: An author-, institution-, and country-level analysis. *Journal of Informetrics, 8*(2), 295–309.

Yang, Z. Q. (2019). *Research on process and driving factors of knowledge complementarity about Industry-university-research institute based on interaction of "Ba."* Chongqing University of Post and Telecommunications. In Chinese.

Yin, Y. Y. (2012). Interactive study of science and technology based on citation relations between papers and patents——Case study of 3D display. *Library and Information Service, 56*(16), 65–70. In Chinese.

Yu, C. M., Gong, Y. T., Zhao, X. L., & An, L. (2017). Collaboration recommendation of finance research based on multi-feature fusion. *Data Analysis and Knowledge Discovery, 1*(8), 39–47. In Chinese.

Zhang, Y., Porter, A. L., Hu, Z. Y., Guo, Y., & Newman, N. C. (2014). 'Term clumping' for technical intelligence: A case study on dye-sensitized solar cells. *Technological Forecasting & Social Change, 85*, 26–39.

Zhang, Y. W., Wu, J., Li, P., Wu, X. J., & Zhou, X. (2015). Partner selection of industry-university-research collaboration based on two-sided matching from the perspective of innovation capacity. *Journal of Jiangsu University of Science and Technology (natural Science Edition), 29*(5), 488–495. In Chinese.

Zhou, T., Lü, L., & Zhang, Y.-C. (2009). Predicting missing links via local information. *The European Physical Journal B, 71*(4), 623–630.