Identifying Technological Trends: A Patent Analysis Method for Technology Forecasting

Patrick D. Kochan[1], Bartholomeo O. Barcelos[2], Alexandre L. Gonçalves[1,2]

[1]Undergraduate Course in Computer Engineering
Federal University of Santa Catarina (UFSC)
88906-072 - Araranguá - SC - Brazil

[2]Graduate Program in Knowledge Engineering and Management
Federal University of Santa Catarina (UFSC)
88040-900 - Florianópolis - SC - Brazil

patrick.kochan@grad.ufsc.br, barthobarcelos@gmail.com, a.l.goncalves@ufsc.br

Abstract. *Patents are extensive and reliable sources of data on technological inventions, serving as the basis for patent analysis tasks. Among these tasks, technology forecasting is essential for research, development and decision-making in organizations. This paper proposes a decision-making support method capable of identifying technological trends. To achieve this, we explore the learning of network representations by applying link prediction algorithms to identify potential trends in the links between technologies. To demonstrate the effectiveness of the proposed method, we conducted experiments in the field of carbon technology. Our link prediction model reached a mean performance of 0.91, considering the ROC-AUC metric.*

1. Introduction

With the advent of Industry 4.0, the organizational environment has become more complex and competitive. This revolution allowed institutions to innovate in their services, processes, and products to establish fundamental differentials for permanence in the market [Ottonicar *et al*. 2018]. For organizations, the ability to innovate stands out as one of the determining factors for achieving superior performance over competitors [Chen *et al*. 2020].

In this sense, patent literature stands out as a broad and valuable source of knowledge and technological inventions for researchers, organizations, and innovation communities [Krestel *et al.* 2021]. These documents offer reliable data and reflect the advances in the technological development of society, which has made patent analysis a vital tool for formulating research, development, and innovation (RD&I) strategies [Zhang and Liu 2020]. According to Zhang and Liu (2020), approximately 80% of the world's technology knowledge can be found in patents. This information is relevant since technologies play an important role in innovation generation, as innovation can be achieved through the convergence of technologies or technological areas [Kim and Bae 2017].

The purpose of the patent analysis is diverse in terms of technical and economic decision-making. The resulting data are used to provide technological capacity indicators, identify novelties, and patent infringements, identify competitors and partners, analyze technology trends, determine the quality of patents and evaluate the innovation process [Zhang and Liu 2020]. From this, priorities in RD&I investment can be defined, as well as the construction of possible strategies for acquiring and merging technologies [Yoon and Lee 2008], [Zhang and Liu 2020].

Nonetheless, the rapid growth in the number of patents represents a major challenge for the recovery and analysis of such information effectively. For this end, the patent analysis describes a group of tasks that can be partially automated. One of these tasks is technology forecasting, in which patents are used to evaluate a given technological scenario, helping researchers to identify technological trends or new technologies [Choi and Song 2018], [Krestel *et al.* 2021]. This task assists in the risk management of emerging innovations and technologies, as well as in the data-driven decision-making process providing tools that have become increasingly necessary for economies, organizations and society. Thus, technology forecasting can benefit institutions by enhancing the identification of business opportunities, reducing risks in the decision-making process and supporting capital investment [Haleem *et al.* 2019], [Krestel *et al* 2021].

In general, this scenario has been promoting an exponential increase in the volume of data resulting from the developments of the digital transformation, such as interconnectivity and technologies such as Big Data and Machine Learning (ML), as well as has evidenced Network Analysis (NA) as a growing research area [Dadu *et al.* 2019], [Ottonicar *et al.* 2018]. As a support to this area, graphs occupy an important position throughout history, due to their broad capacity to characterize real-world problems. Their rich representation of the data reveals the relationships between entities in a simple way. The entities are modeled as vertices, and their relationships are arranged as edges [Wang *et al.* 2019]. This structure has been used to represent social media, connections between web pages, geographic maps, and numerous other possibilities of interconnection between data [Gupta *et al*. 2021].

Also, in the case of the field of network analysis, link prediction is a task that has been highlighted in recent years and covers the most diverse application domains, such as the recommendation of friends on social media, protein-protein interactions in the field of bioinformatics and product recommendation systems in e-commerce [Kumar *et al.* 2020]. To this end, different techniques are used to measure the probability of two nodes of a network connecting [Wu *et al.* 2021].

Thus, this scientific research has its relevance by pointing out a possible method of identifying technological trends from the concept of link prediction and by contributing to supporting the strategic decision-making process in organizational environments. Also, this study's results are expected to help promote new research on related topics.

2. Theoretical Background

2.1. Patent Analysis

With the rapid economic development of society, high technology plays a crucial role in the development of companies, countries, and society as a whole [Abbas *et al.* 2014]. This produces the need to evaluate technologies before adopting them in a project, avoiding use violations and ensuring that they have good RD&I prospects. Since patents are reliable sources of knowledge about technical progress and innovative activity and constitute a reliable resource that reflects advances in technological development, patent analysis has been considered a vital tool for the formulation of strategies in the RD&I area [Zhang and Liu 2020].

Abbas *et al.* (2014) state that the increasing volume of technical data related to technological inventions has made the analysis task extremely laborious. The number of existing patents comprises millions of documents spread over different databases available from web-based sources (open data) or private sources. The most popular repositories for these documents are the *United States Patent and Trademark Office®* (USPTO), *European Patent Office®* (EPO) and *Japan Patent Office®* (JPO). According to Yoon and Lee (2008), relying solely on the knowledge and skills of specialists to analyze patents has become impracticable and, therefore, the use of auxiliary tools becomes imperative. Therefore, computational tools can be used to reduce not only the extensive manual analysis of patents by specialists but also accelerate the steps of the process: extraction of patents from databases; extraction of patent information; and the analysis of extracted information.

According to Krestel *et al.* (2021) and Lupu *et al.* (2017), the patent study requires the in-depth work of specialists and the evaluation of a huge volume of data. All the work involving patent analysis can be divided into tasks that can be partially automated. Krestel *et al.* (2021) separates these tasks into eight main ones, including patent classification and technology forecasting tasks, the latter being the purpose of this work, with the aim of assisting to recognize technological trends.

2.2. Technology Forecasting

Technology forecasting is among the most popular patent analysis tasks. Its objective is to identify technological opportunities and generate valuable insights for governments, companies, and decision-makers in order to guarantee support for capital investment [Kim *et al.* 2019] [Krestel *et al.* 2021]. Also, this task has been recognized as an essential step in the RD&I process [Kim and Bae 2017]. In this context, patents emerge as a valuable resource for forecasting and decisions about which technologies to adopt, as they provide updated and reliable knowledge for identifying technological trends [Altuntas *et al.* 2015].

Lenz (1962), one of the pioneers of technology forecasting, defined it as the prediction of inventions, characteristics or performance of a machine serving some useful purpose and pointed out that the qualities sought for prediction methods are explicitness, quantitative expression and reproducibility of results. For Cho and Daim (2013), technology forecasting is the ability to analyze and evaluate the performance parameters of a product through probability statements with a relatively high confidence level, capturing opportunities and threats of technological changes to provide valuable information for decision-making in RD&I.

2.3. Link Prediction

According to Barabási and Pósfai (2016), graph theory is the historical mathematical background of modern Network Science (NS) including link prediction. The goal of link prediction is to forecast missing interactions that may occur in the future in an evolving network. This task is one of the fundamental problems of network analysis and is deeply associated with recommendation systems [Amara *et al.* 2021]. The first attempts to implement link prediction used heuristic seeking to capture and exploit a few structural information of the network, such as the common neighbor algorithms, the Jaccard similarity coefficient, and the Adamic-Adar index [Malek *et al.* 2021].

Kumar *et al.* (2020) understand the link prediction as follows: given a non-targeted graph $G(V, E)$ where V characterizes the set of vertices and E the set of edges, this graph has a universe set U which has a total of $\frac{n.(n-1)}{2}$links, where $n = V$. The non-existent links are determined by $U - E$, and some of these links may emerge in the near future. Finding the missing links is the goal of the link prediction. Formally, Liben-Nowell and Kleinberg (2003) defined the problem of link prediction as: a graph $G_{t0-t1}(V, E)$ represents an instant of the network during the time interval $[t0, t1]$ and E_{t0-t1} represents the set of links present at that time. The task of predicting links is finding the set $E_{t'0-t'1}$ over a time interval $[t'0, t'1]$, where $[t'0, t'1]$ $[t0, t1]$. Figure 1 graphically represents the task, where, for example, among three possible links, only one was predicted as the more likely to occur.

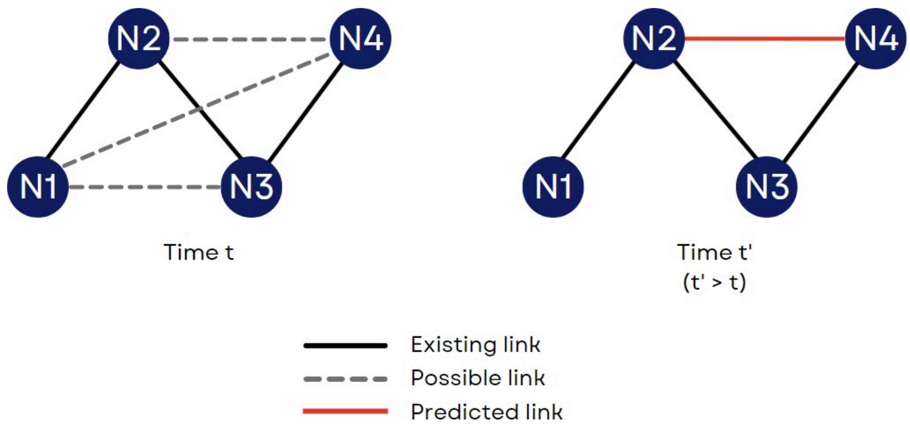

Figure 1. Link prediction task

3. Related Works

Through searches in the scientific literature, studies were selected in order to identify methods and approaches for patent analysis used in technology forecasting employing prediction techniques and trend identification. The research was conducted in the academic articles databases ACM Digital Library®, ScienceDirect®, Scopus® and Web of Science®. Scientific papers in English with combinations of the terms: “Patent*”, “Link Prediction”, “Forecasting or Prediction”, “Emerging Technolog*”, were considered, where “*” represents the possible variations of the terms. Of the total of 184 resulting studies, 30 presented an appropriate title for the purpose of the search. Their abstracts were then read in full, and 13 articles were selected for reading the introduction. Finally, the 9 works related to this research's theme that have passed through all the exclusion criteria are summarized below, considering the dimensions, techniques and methods used, contributions, results obtained and conclusions.

In the evaluated works, a variety of techniques and methods for analyzing patent data were used. Ma, Pan and Su (2022) and Lee *et al.* (2021) applied link prediction techniques to explore technological opportunities in co-occurrence networks, the former focusing on manual codes (MCs) and the latter on *F*-term classifications. Park and Yoon (2018) also found that link prediction could be a powerful tool in order to identify technological convergence opportunities. Choi and Song (2018) and Kim *et al.* (2019) used the Latent Dirichlet Allocation (LDA) topic modeling technique to extract thematic trends from patent data. Also, this method was essential for identifying emerging technologies and technology gaps in the context of wireless power transfer, as demonstrated by Kim *et al.* (2019). Meanwhile, a deep learning approach using a Generative Adversarial Network (GAN) was adopted by Zhou *et al.* (2020) to enhance the dataset aiming to increase the accuracy of emerging technology predictions. The use of machine learning was further explored by *Lee et al.* (2018). The authors applied a multilayer feed-forward neural network to classify patents based on predicted future citations, underlining the growing role of artificial intelligence in patent analysis.

The contributions of these studies to the field of technology forecasting are varied. Ma, Pan and Su (2022) provided insights on the dynamics of technological knowledge networks through a method for expanding the technological boundaries of organizations. Choi and Song (2018) and Kim *et al.* (2019) contributed to the understanding of technological trends and gaps, with implications for both industry-specific forecasting and broader technological developments. Zhou *et al.* (2020) advanced the application of deep learning in forecasting by demonstrating its effectiveness in scenarios with limited data. Jeong *et al.* (2021) introduced a framework for business diversification extending patent analysis beyond technology forecasting to include strategic business decisions. Kim and Bae (2017) and Lee *et al.* (2021) were concerned with identifying promising technologies using cluster analysis and link prediction to guide research and development (R&D) efforts.

The results obtained in these studies demonstrate the effectiveness of the methods used. In the work by Ma, Pan and Su (2022), a minimum ROC-AUC of approximately 85% was achieved with the accuracy metric ranging from 27.81% to 50%, indicating a suitable approach for assessing innovation activities. Choi and Song (2018) successfully categorized topics in the field of logistics into four groups of trends in patent activities, while Kim *et al.* (2019) identified one emerging technology area and two existing technology gaps in wireless power transfer. Zhou *et al.* (2020) achieved, even with data limitations, over 77% accuracy in predicting emerging technologies validating the potential of deep learning in this domain. Jeong *et al.* (2021) demonstrated a high accuracy of 81.87% in predicting business diversification opportunities. Kim and Bae (2017), on the other hand, concluded that telemedicine was a promising area in the health sector using both classification and clustering techniques. Finally, Lee *et al.* (2021) were able to accurately predict 39 out of 100 technological opportunities that corresponded to real R&D results, demonstrating the practical relevance of their link prediction approach.

The conclusions drawn from the discussed papers indicate the potential of patent analysis in technology forecasting. Ma, Pan and Su (2022) emphasized the value of their approach based on link prediction in expanding the technological frontiers of organizations. Choi and Song (2018) highlighted the benefit of LDA in capturing specific trends in the logistics sector, while Kim *et al.* (2019) and Zhou *et al.* (2020) demonstrated the accuracy and validation of their methods in identifying gaps and forecasting emerging technologies, respectively. Jeong *et al.* (2021) and Park and Yoon (2018) used patent analysis focusing on business strategy and technological convergence pointing to new possibilities for research and application. The studies by Kim and Bae (2017), Lee *et al.* (2018) and Lee *et al.* (2021) emphasized the use of machine learning techniques, such as clustering, as well as the use of neural networks in predicting promising technologies and early-stage opportunities.

In summary, these studies promote advances in the field of patent-based technology forecasting through different methods and techniques, with important theoretical and practical contributions. The conclusions underline the importance of continuous innovation through analytical approaches aimed at harnessing the potential provided by patent data in forecasting and modeling future technological scenarios.

4. Proposed Method

This paper presents a patent analysis method for technology forecasting to support strategic decision-making in organizational environments. To achieve this goal, it uses network representation learning techniques and link prediction in the context of patent analysis. Figure 2 provides an overview of the method and the five steps that compose it. Next, each of the steps is described in detail.

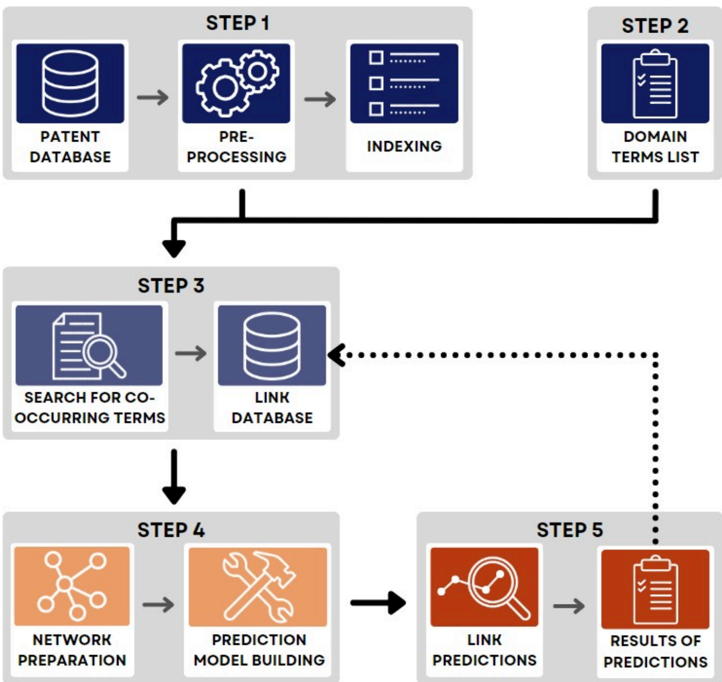

Figure 2. Overall process of the proposed method

Step 1: Collecting, Pre-processing, and Indexing Data

The first step of the proposed method is related to the collection and indexing of patent documents. For example, each patent document is processed by removing punctuations and considering the most relevant fields, such as title, abstract, and year. Afterwards, patents are indexed in a document-oriented database. This type of storage facilitates the handling and searching of the data and, due to the flexibility of the structure, allows the evolution of the data according to the need for development. As a result of this step, there is an indexed patent database where it is possible to perform searches for keywords.

Step 2: Defining the domain and terms of interest

For the second step, an area of interest must initially be chosen. This is the key element of this step. The full list of technologies and concepts formed from it will be part of certain analysis. Next, data sources that present terms intrinsically related to the topic of interest are located. The terms capable of describing the domain are then collected and integrated to produce a list that serves as the basis for the target study.

Step 3: Populating the link database

The third step is responsible for joining the first two results to form the link database. To this end, queries are conducted in the indexed patent database based on the terms of the list obtained in the second step. First, all the different possibilities of relations between the terms are generated. Then, the queries are carried out so that joint frequencies between them are stored year by year. Finally, the number of co-occurrences, the year, and the co-occurring terms in the source-target format are stored in a relational database.

Step 4: Building the prediction model

Next, step four prepares the network and builds the link prediction model. Initially, the link database is used to generate the network itself. Then, a set of positive and negative samples is formed that will feed the prediction model. At this point, some links are removed from the network to compose the positive samples, allowing the model to be able to learn how these true connections occurred. This is accomplished through the random removal of links and the verification that the number of connected components of the network remains the same. On the other hand, negative samples are composed of all the non-existent links of the network, obtained through an adjacent matrix. The learning of network representation is then applied, and, lastly, the prediction model is generated from the representations of the links and the set of samples.

Step 5: Predicting and presenting the results

Finally, the link predictions are made through the built model in the fifth step. A minimum acceptable probability is defined to consider a prediction to be true. As a result, we obtain a list of technologies or concepts linked to the chosen domain that will probably connect in the near future. This list is then used to check whether the predictions occurred as new patents are indexed, and new relationships are included in the database.

5. Experiment

To demonstrate the feasibility of the proposed method, an experiment was conducted. The data set was constituted from patents obtained from the United States Patent and Trademark Office® (USPTO) between 2006 to 2015 published in the study by [Li *et al.* 2018]. This data set, named USPTO-2M®, is used in tests and validation of tasks involving patents, especially the classification task, and has 2,000,147 documents. It is composed of utility patents, that is, patents that cover the creation of a new or improved product, process, or machine, meeting the objective of this work.

In this experiment, the theme was defined by the term “carbon”, as it has been the subject of a large volume of scientific research. A brief query on the academic publications search engine Semantic Scholar® revealed that, over the past ten years, approximately two million results presented the term. Predictions, in short, medium, and long-term scenarios, were considered to evaluate the proposed method, considering periods of one, three, and five years.

5.1. Application of the proposed method

To constitute the patent database of the experiment, the USPTO-2M® data set composed of patents from 2006 to 2015 was considered. The patent processing was then initiated, removing the punctuation from the “title”, and “abstract” fields. The year field is also a fundamental characteristic, as it allows filters in future steps of evaluating the method. From this, each document was indexed in a document database in order to perform searches in full text by keywords.

As pointed out in the presentation of the experiment, “carbon” was the domain chosen due to its notoriety in recent research. A list of about 500 terms was obtained from several online glossaries consulted. Of these, 148 terms that co-occurred most among the glossaries were considered to compose a domain list. Some terms are “Carbon Footprint”, “Net Zero Carbon Business”, “Biofuel”, “Fugitive Emissions”, ‘Carbon Nanotube”, “Global Warming”, “Recycling”, and “Carbon Fiber”.

From the list of domain terms, all possible distinct combinations of relationships between the terms were generated. Then, through a document database, the search for co-occurring terms, that is, pairs of terms, was carried out. In all, 3719 co-occurrences were generated between 79 terms of the “carbon” domain list when considering the entire experiment period. From the results of the queries, the link database supported by a MySQL® database was generated. The frequencies of individual term mentions and the co-occurring terms in the source-target format were stored, year by year, for the entire interval from 2006 to 2015.

The training period was first defined to construct the prediction model so that the evaluation could occur. The period from 2006 to 2010 was selected so that all the scenarios mentioned in the beginning of this section could be evaluated in relation to the year in which the prediction model was trained. From the given year, a network was built through the NetworkX® library in Python® programming language.

The set of positive and negative samples was then formed from the annual networks. Table 1 presents examples of these samples collected from the 2006 network, where the value “0” in the link column indicates the absence of the link between the source-target nodes and “1” the presence. To carry out the network representation learning, the Node2vec® algorithm was applied, and the vector representations of the links were obtained. Then, the link representation set and the sample set were separated into two representation-target pairs: one pair for the test, with 30% of the data, and another for training, with the rest of the data. Finally, the prediction model for the given year was generated based on the logistic regression through the Scikit-learn®, open source machine learning library.

In the final step, the trained model was used to calculate predictions for all potential network connections, resulting in a list of terms (technologies and concepts) that may be connected in the following years.

Table 1. Examples of negative and positive samples

Sample type	Source node	Target node	Link
Negative	fugitive emissions carbon black biofuel	carbon fiber bioprocessing recycling	0 0 0
Positive	carbon composite recycling biomass	carbon material biogas atmosphere	1 1 1

5.2. Results and Analysis

The analysis of the proposed method occurred as follows: i) from each model trained in the interval from 2006 to 2010, 100 runs were performed for three prediction thresholds, with 50%, 70%, and 85% probability of link in three time scenarios, short (one year), medium (three years), and long (five years); ii) the mean of the evaluation metrics of the model and the predictions in relation to the 100 runs was obtained to ensure fidelity of the results.

The metric used to evaluate the prediction model was the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). This metric is widely used as a performance indicator for link prediction and binary classification problems [Lee *et al.* 2021]. The ROC-AUC metric calculates the area under the ROC curve, that is, a probability curve representing the true positive rate versus the false positive rate at different classification thresholds. In general, it simplifies the ROC curve analysis by aggregating all its thresholds and summarizing it to a single value. Basically, the closer to 1 to ROC-AUC, the better the classifier’s performance by distinguishing between the links that occurred and those that did not. When considering the ROC-AUC of all models trained in the interval from 2006 to 2010 (Table 2) the prediction model was able to achieve a mean performance of 0.91, which demonstrates that the classifier is highly capable of distinguishing between the positive and negative link samples considering the test set.

After the initial evaluation of the prediction model, the evaluation of the proposed method was initiated from the scenario of the experiment to determine its real performance. For this, the steps described in subsection 4.1 were considered and the results are summarized in Table 3 and Table 4.

Table 2. ROC-AUC obtained by each trained prediction model

Year	2006	2007	2008	2009	2010
ROC-AUC	0.89	0.91	0.91	0.93	0.91

In order to facilitate the visualization of the data and monitoring of the analyses, Table 3 categorizes the percentage of correct predictions made in: i) poor, [0%, 33%), highlighted in red; ii) satisfactory, [33%, 66%), highlighted in yellow; and iii) good, [66%, 100%], highlighted in green. Furthermore, for each year, the table is divided into three sections, representing the total predictions considering thresholds greater than or equal to 50%, 70% or 85%. That is, the 50% threshold indicates that only the predictions of links in which the probability is greater than or equal to 0.5 should be considered to be evaluated in the short, medium, and long-term scenarios. The number of predictions is indicated in the column "Average Amount of Predictions", since 100 runs were conducted. The success percentage rates appear in the "Success Percentage" columns.

From Table 3, it was found that the scenario that best performed in terms of the success percentage was the long-term scenario, regardless of the training year of the model and the prediction threshold. This result shows that the predictions made have, in fact, the potential to occur in the future.

Considering the prediction threshold and the entire range from 2006 to 2010, the 85% threshold provided better results in 4 of the 5 years evaluated. The exception was 2010, where the highest success rate was 24% with the threshold of 70%, very close to the accuracy of 23.92% obtained by the 85% threshold. Comprehensively, except for the example mentioned above, the interpretation of the results is since the number of predictions made decreases as the probability of the link occurring increases, consequently increasing the percentage of correct predictions. When considering the entire analysis period, the average amount of predictions made was 188 for the 50% threshold, 52 for the 70% threshold, and 11 for the 85% threshold, which directly influences the success rates.

Another relevant factor in the discussion is the network complexity. When comparing the data in Table 3 and Table 4, it is observed that the percentage of correct predictions decreases year by year, and, on the other hand, the number of nodes and links present in the network increases. This behavior shows that the dynamic aspect of networks has a major impact on the task of technology forecasting. The evolution trend of the network assumes that it gets bigger and more complex over time. This makes it more difficult to assign representation and predict which links will occur in practice, as the inclusion of new entities implies the emergence of numerous new possible links. From this perspective, it is necessary to increase the network representation learning capacity and the robustness of the predictive model to obtain better results in complex networks.

Table 3. Metrics obtained from the predictions made from 2006 to 2010

Scenario (term)	50% Average Amount of Predictions	50% Success Percentage	70% Average Amount of Predictions	70% Success Percentage	85% Average Amount of Predictions	85% Success Percentage
2006						
Short	173	16,71%	51	22,92%	11	25,26%
Medium	172	36,76%	51	52,60%	12	68,43%
Long	173	50,27%	51	65,58%	11	78,06%
2007						
Short	189	11,43%	48	19,50%	6	44,37%
Medium	188	32,24%	47	44,74%	6	57,26%
Long	188	40,67%	48	55,21%	7	76,36%
2008						
Short	183	8,61%	52	11,90%	7	17,24%
Medium	183	26,64%	52	33,59%	7	38,98%
Long	184	33,56%	53	42,44%	7	50,08%
2009						
Short	207	12,46%	77	15,92%	18	14,87%
Medium	205	22,00%	77	27,88%	18	29,16%
Long	205	32,00%	75	40,66%	19	41,46%
2010						
Short	234	6,85%	91	8,72%	25	8,58%
Medium	234	14,31%	92	17,13%	25	14,97%
Long	238	20,07%	90	24,00%	25	23,92%

Considering the long-term scenario and the 85% prediction threshold, the proposed method reached a mean success percentage of 53.98% for the period analyzed in the experiment. In networks of lower complexity, that is, with a lower number of nodes and consequently possible links, the method obtained a better performance, reaching up to 78.06% of success in the predictions made. Therefore, the built prediction model can fulfill its task, promoting evidence, through the prediction of links, of possible technological trends. From the application perspective, the proposed method can support the strategic decision-making process since its impact in terms of competitive intelligence and construction of business strategies stands out given the accuracy of the predictions.

Table 4. Networks information

Year	Number of nodes	Number of links
2006	50	261
2007	52	325
2008	54	359
2009	59	397
2010	63	461

5. Conclusions and Future Work

This work presented a patent analysis method towards the technology forecasting task to support strategic decision-making in organizational environments. For this purpose, the work employed network representation learning based on shallow neural networks and link prediction taking into account the similarity with deep learning methods in terms of predictive power, but using less computational resources. It was decided to adopt an approach based on neural networks to guarantee the proposed method neutrality in terms of domain since the technique can adapt to the application context.

In this research, 2.000.147 patent documents from the USPTO-2M® data set were collected for an experiment on the “carbon” domain. The evaluations of the proposed method considered predictions in short (one year), medium (three years), and long (five years) term scenarios under prediction thresholds of 50%, 70%, and 85%. The results showed that, regardless of the threshold and the training year of the model, the best prediction scenario was the long-term one. On average, the long-term success rate of the predictions made under the 85% threshold was 53.98%. Besides, the link prediction model reached a mean ROC-AUC of 0.91. Thus, the proposed method is feasible for the presented conjuncture.

Despite the contributions of the proposed method, this study is subject to limitations that require future work. First, in order to deal with more complex networks, it is essential to work on the learning capacity of network representation and the robustness of the predictive model. Second, even if the organization defines the domain and list of technologies according to its interests, the predictions do not take into account the specific objectives and the technological capacity of the institution. Also, the terms that describe the domain and the patent database must be updated periodically to reflect the latest technologies, which may cause an undesired delay in identifying opportunities.

References

Abbas, A., Zhang, L. and Khan, S. U. (2014) “A literature review on the state-of-the-art in patent analysis”. In *World Patent Information,* n. 37, pages 3–13.

Altuntas, S., Dereli, T. and Kusiak, A. (2015) “Forecasting technology success based on patent data”. In *Technological Forecasting and Social Change,* n. 96, pages 202–214.

Amara, A., Taieb, M. A. H. and Aouicha, M. B. (2021) "Network representation learning systematic review: Ancestors and current development state". In *Machine Learning with Applications,* n. 6, pages 100-130.

Barabási, A. L. and Pósfai, M. (2016) "Network science". Cambridge University Press, Cambridge, URL: http://networksciencebook.com/

Chen, Q., Wang, C. H. and Huang, S. Z. (2020) "Effects of organizational innovation and technological innovation capabilities on firm performance: evidence from firms in China's pearl river delta". In *Asia Pacific Business Review* vol. 26, n. 1, pages 72–96.

Cho, Y. and Daim, T. (2013) "Technology forecasting methods". In: Daim T, Oliver T, Kim J (eds) Research and Technology Management in the Electricity Industry: Methods, Tools and Case Studies. Springer London, London, pages 67-112.

Choi, D. and Song, B. (2018) 'Exploring technological trends in logistics: Topic modeling-based patent analysis". In *Sustainability,* vol 10, n. 8., pages 1-26.

Dadu, A., Kumar, A., Shakya, H., *et al.* (2019) "A study of link prediction using deep learning". In *Communications in Computer and Information Science*, vol 955, pages 377–385.

Gupta, A., Matta, P. and Pant, B. (2021) "Graph neural network: Current state of art, challenges and applications". In *Materials Today: Proceedings,* vol. 46, n. 20, pages 10927–10932.

Haleem, A., Mannan, B., Luthra, S., *et al.* (2019) "Technology forecasting (TF) and technology assessment (TA) methodologies: a conceptual review". In *Benchmarking: An International Journal*, vol. 26, n. 1, pages 48–72.

Jeong, B., Ko, N., Son, C., *et al.* (2021) "Trademark-based framework to uncover business diversification opportunities: Application of deep link prediction and competitive intelligence analysis". In *Computers in Industry*, vol. 124, p. 103356.

Kim, G. and Bae, J. (2017) "A novel approach to forecast promising technology through patent analysis". In *Technological Forecasting and Social Change*, vol. 117, pages 228–237.

Kim, K. H., Han, Y. J., Lee, S., *et al.* (2019) "Text mining for patent analysis to forecast emerging technologies in wireless power transfer". In *Sustainability*, vol. 11, n. 22., p. 6240.

Krestel, R., Chikkamath, R., Hewel, C., *et al.* (2021) "A survey on deep learning for patent analysis". In *World Patent Information*, vol. 65, p. 102035.

Kumar, A., Singh, S. S., Singh, K., *et al.* (2020) "Link prediction techniques, applications, and performance: A survey". In *Physica A: Statistical Mechanics and its Applications*, vol. 553, p. 124289.

Lee, C., Kwon, O., Kim, M., *et al.* (2018) "Early identification of emerging technologies: A machine learning approach using multiple patent indicators". In *Technological Forecasting and Social Change*, vol. 127, pages 291–303.

Lee, J., Ko, N., Yoon, J., *et al*. (2021) "An approach for discovering firm-specific technology opportunities: Application of link prediction to F-term networks". In *Technological Forecasting and Social Change,* vol. 168, p. 120746.

Lenz, R. C. (1962) "Technological forecasting". Report ASDTDR 62-414. *Aeronautical Systems Divisions, Wright-Patterson Air Force Base, Ohio*.

Li, S., Hu, J., Cui, Y., *et al.* (2018) "Deep patent: patent classification with convolutional neural networks and word embedding". In *Scientometrics,* vol. 117, n. 2, pages 721–744.

Liben-Nowell, D., and Kleinberg, J. (2003) "The link prediction problem for social networks. In: Proceedings of the Twelfth International Conference on Information and Knowledge Management". Association for Computing Machinery, New York, NY, USA, CIKM'03, pages 556–559.

Lupu, M., Fujii, A., Oard, D. W., *et al.* (2017) "Patent-related tasks at NTCIR". In: Current Challenges in Patent Information Retrieval. Edited by Lupu, M., Mayer, K., Kando, N., *et al*. Springer Berlin Heidelberg, Berlin, Heidelberg, pages 77– 111.

Ma, J., Pan, Y., and Su, C. Y. (2022) "Organization-oriented technology opportunities analysis based on predicting patent networks: a case of Alzheimer's disease". In: *Scientometrics,* vol. 127, n. 9, pages 5497–5517.

Malek, M., Chehreghani, M. H., Nazerfard, E., *et al.* (2021) "Shallow node representation learning using centrality indices". In: 2021 IEEE International Conference on Big Data (Big Data), pages 5209–5214.

Ottonicar, S., Valentim, M., Mosconi, E. (2018) "A competitive intelligence model based on information literacy: Organizational competitiveness in the context of the 4th industrial revolution". In *Journal of Intelligence Studies in Business,* vol. 8, pages 55–65.

Park I, Yoon, B (2018) "Technological opportunity discovery for technological convergence based on the prediction of technology knowledge flow in a citation network". In *Journal of Informetrics*. vol. 12, n. 4, pages 1199–1222.

Wang, W., Wu, L., Huang, Y., *et al.* (2019) 'Link prediction based on deep convolutional neural network". In *Information,* vol. 10, n. 5. pages 1-17.

Wu, W., Li, B., Luo, C., *et al.* (2021) "Hashing-accelerated graph neural networks for link prediction". In: Proceedings of the Web Conference 2021. Association for Computing Machinery, New York, NY, USA, WWW'21, pages 2910–2920.

Yoon, B. and Lee, S. (2008) "Patent analysis for technology forecasting: Sector-specific applications". In: 2008 IEEE International Engineering Management Conference, pages 1–5.

Zhang, L. and Liu, Z. (2020) "Research on technology prospect risk of high-tech projects based on patent analysis". In *Plos One*, vol. 15, n. 10, pages 1–19.

Zhou, Y., Dong, F., Liu, Y., *et al.* (2020) "Forecasting emerging technologies using data augmentation and deep learning". In *Scientometrics*, vol. 123, n. 1, pages 1–29.