Evaluating and ranking patents with multiple criteria: How many criteria are required to find the most promising patents?

Nikolaos Ploskas, Tong Zhang, Nikolaos V. Sahinidis

Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Flor Castillo, Krishnan Sankaranarayanan

SABIC Americas, Sugar Land TX, 77478, USA

Abstract

Patents contain a wealth of information about technological progress and market trends. Many existing techniques for patent assessment rely on citation analysis. Despite its importance, citation analysis alone is not adequate to identify all important patents for a given topic. We propose the simultaneous use of eight criteria for patent ranking and evaluation. Additionally, we investigate computationally the effect on ranking quality when fewer than eight criteria are utilized. Contrary to previous approaches, the proposed methodology does not require expert opinions to weigh the different criteria and evaluate the patents. The solution of an intuitive linear optimization problem provides optimal weights for the proposed criteria. These weights are subsequently utilized in a systematic multicriteria methodology for patent ranking. The proposed methodology has been implemented in a web-based decision support system and has been validated in the context of identifying the most important patents for the production of twenty-two chemicals.

Keywords: Decision support systems, Multiple criteria analysis, Text analytics, Patent rankings, Optimization

*Corresponding author
Email address: sahinidis@cmu.edu (Nikolaos V. Sahinidis)

1. Introduction

Patents contain a wealth of information about technological progress and market trends. Although patent documents have a well-defined format, they are lengthy and include many technical terms, which requires significant effort to analyze. Hence, an entire research area, called patent mining or patent analysis, aims to assist patent analysts and policy makers in finding, processing, and analyzing patents. Patent analysis can reveal market trends and novel industrial solutions that can lead to investment decisions [1].

The basic tasks that a patent mining system performs are [2]:

- Patent search and retrieval: the system searches for relevant patents from patent databases according to a keyword/phrase. Natural language processing and data mining methods have been used to improve the relevance of the returned results for a query [3, 4, 5, 6, 7, 8, 9, 10].

- Patent visualization: patents are presented in ways that help analysts understand key relationships and concepts. Graph theory, network analysis, and text mining methods have been applied to visualize patents that contain both structured and unstructured data [11, 12, 13, 14, 15, 16, 17].

- Patent evaluation: patents are evaluated according to their importance or potential. This challenging but important task can help decision makers invest in new novel industrial solutions. Natural language processing and data mining methods have been used for patent evaluation [18, 19, 20] (a detailed literature review is provided in Section 2). The most frequently used technique is citation analysis [21, 22]. Although citation analysis can reveal a seminal discovery through the number of citations that a patent received, it may miss important recently issued patents.

In this paper, we propose a new methodology for patent evaluation and ranking. Our methodology takes into account several pieces of information about patents, including citation counts, pagerank, and the existence of clusters in citation graphs. The relative importance of the proposed criteria is determined through the solution of a linear optimization model that provides weights that satisfy a number of intuitive constraints amongst the proposed criteria. These weights constitute the main ingredients of a systematic patent evaluation and ranking method that utilizes the multicriteria methodology TOPSIS [23]. The proposed approach represents the first multiple criteria approach to patent evaluation and ranking that does not require estimations from decision makers. Instead, we rely on common sense constraints that are easier to argue and negotiate between experts. Additionally, we investigate computationally the impact of the number of criteria used to evaluate patents. These computations reveal that at least five criteria are needed to obtain somewhat similar rankings and that none of the eight proposed criteria is redundant in the sense that their inclusion in the model leads to different rankings.

The remainder of this paper is organized as follows. In Section 2, we present a literature review on patent evaluation and ranking techniques. Section 3 details the proposed methodology. Section 4 describes a computational implementation of the proposed approach in a web-based decision support system. Section 5 presents computational experiments, including finding the most important patents for the production processes of twenty-two chemicals. Conclusions from this research are presented in Section 6.

2. Related work

Patent evaluation techniques aim to support decision makers by assessing the quality of patents. Most existing works have relied on citations to rank patents. This includes forward citations, i.e., citations received by a patent from patents granted at a subsequent point in time, and backward citations, i.e., citations given by a patent to patents granted at an earlier time. Previous studies [24, 25, 26, 27] have shown that patent citations can be used to evaluate the novelty and importance of a patent.

Hasan et al. [28] presented a patent ranking software, named COA (Claim Originality Analysis), which rates a patent based on its value and the impact of the important phrases that appear in the Claims section of a patent. The metrics used for patent ranking by the authors [28] were the patent citation count (the number of citations that a patent receives from other patents), the patent status (whether the patent is still maintained by its assignee), and confidential attorney ratings. Computational experiments with this approach were performed on IBM patent portfolios. Jin et al. [29] proposed a patent information network model in order to assess the value of a patent. The metrics used by these authors were the publication year, filed year, number of investors, number of assignees, number of claims, number of US classes, number of IPC codes, US patent kind, number of forward citations, number of backward citations, number of other citations, and number of total citations. Jin et al. [29] performed an experiment on patents from companies and organizations with large patent portfolios from a variety of fields. Liu et al. [20] introduced a latent graphical model to infer patent quality using natural language processing techniques. The metrics used in this work were the number of forward citations, court decisions (ruled as valid or invalid), and reexamination records. Liu et al. [20] performed an experiment on a set of approximately $12,000$ randomly selected patents and a set of 351 patents with decisions by the Federal Circuit Court.

Oh et al. [30] proposed a weighted citation method to evaluate and rank patents based on four different types of citations received by a given patent: (i) citations received from patents filed by the same assignee, (ii) citations received from patents filed by different assignees, (iii) citations received from patents from the same technology domain, and (iv) citations received from patents from different technology domains. An experiment was conducted on the National Bureau of Economic Research patent data project. Hu et al. [19] introduced a topic-based temporal mining approach to quantify patents novelty and influence. They extracted topics from the title, abstract, claims, and detailed description sections of patents and assessed patents novelty and influence by analyzing activity of a patent's topic over time. The metrics used in this work were the number of forward citations and the patent maintenance status. Their experiments were performed on a dataset containing 82,648 U.S. patents from 108 large petroleum companies. Barbazza et al. [31] and Collan et al. [32] proposed a multi-expert system for ranking patents using the TOPSIS and the Analytic Hierarchy Process. The ranking metrics used by these authors were strategic fit, technical quality, licensing potential, ability to disturb competitors activities, ability to open new markets, and ability to protect the company's own activity. Barbazza et al. [31] and Collan et al. [32] performed an experiment on a randomly generated numerical example and required the participation of experts to determine how to weigh different metrics.

Lawryshyn et al. [33] developed a decision support system with a unique real options framework to value and rank patents. They relied on a group of experts to provide estimates for cash-flows and costs of patent projects. An experiment was performed on a randomly generated numerical example. Oh et al. [34] proposed an approach to assess the value of a patent by exploring technologically relevant prior patents as a supplement to backward citations. They accounted for the number of claims, figures, inventors, assignees, foreign references, other references, USPC codes, and IPC codes. Oh et al. [34] conducted an experiment on four million U.S. patent documents granted from 1980 to 2012. Wang and Hsieh [35] proposed a fuzzy multiple criteria decision making survey for patent evaluation. They ran a factor analysis to extract 10 independent criteria for evaluating patents: (i) business potential, (ii) patent quality value, (iii) revenue creation of patent application in relevant industry, (iv) innovativeness of technology, (v) residual life cycle of patent, (vi) competitiveness of technology, (vii) new products and/or processes initiated in relevant industry, (viii) organization growth, (ix) new products initiated in non-relevant industry, and (x) new processes initiated in non-relevant industry. Wang and Hsieh [35] performed an experiment on 4,346 patents of the Industrial Technology Research Institute. Zhang et al. [36] proposed an entropy-based weighting model for ranking patent potential in technological innovation. The ranking metrics used by these authors were: (i) the number of inventors, (ii) the number of patent families, (iii) the number of legal transactions, (iv) the number of claims, (v) the number of patent references, (vi) the number of non-patent references, (vii) the number of citations, (viii) the number of IPCs, (ix) the number of terms, (x) time gap, and (xi) the number of assignees. Their experiments were performed on a dataset containing $28,509$ U.S. patents.

Table 1 summarizes the metrics that have been used to evaluate and rank patents.

Table 1: Metrics used to evaluate and rank patents

Publication	**Metrics**
Hasan et al. [28]	Patent citation count
	Patent status
	Confidential attorney ratings
Jin et al. [29]	Publication year
	Filing year
	Number of investors
	Number of assignees
	Number of claims
	Number of US classes
	Number of IPC codes
	US patent kind
	Number of forward citations
	Number of backward citations
	Number of other citations
	Number of total citations
Liu et al. [20]	Number of forward citations
	Court decisions
	Reexamination records

Oh et al. [30]	Citations received from patents filed by the same assignee Citations received from patents filed by different assignees Citations received from patents from the same technology domain Citations received from patents from different technology domains
Hu et al. [19]	Topic activity Number of forward citations Patent status
Barbazza et al. [31]	Strategic fit Technical quality Licensing potential Ability to disturb competitors activities Ability to open new markets Ability to protect the company's own activity
Lawryshyn et al. [33]	Estimations from a group of experts for low, medium and high cash-flows for patent projects and patent costs
Oh et al. [34]	Number of claims Number of figures Number of inventors Number of assignees Number of foreign references Number of other references Number of USPC code Number of IPC codes

Wang and Hsieh [35]	Business potential
	Patent quality value
	Revenue creation of patent application in relevant industry
	Innovativeness of technology
	Residual life cycle of patent
	Competitiveness of technology
	New products and/or processes initiated in relevant industry
	Organization growth
	New products initiated in non-relevant industry
	New processes initiated in non-relevant industry
Zhang et al. [36]	Number of inventors
	Number of patent families
	Number of legal transactions
	Number of claims
	Number of patent references
	Number of non-patent references
	Number of citations
	Number of IPCs
	Number of terms
	Time gap
	Number of assignees

Almost all of the aforementioned studies perform citation analysis to assess the quality and/or importance of a patent. Although citation analysis can reveal an important discovery through the number of citations that a patent received, it is not adequate to identify important newly or recently issued patents. Moreover, all multiple criteria decision making methods that have been applied for patent evaluation and ranking require estimations from a group of experts and cannot be used for automatically ranking hundreds or thousands of patents. In this work, we present the first multiple criteria decision making method for ranking patents without requiring user intervention, such as estimations from decision-makers.

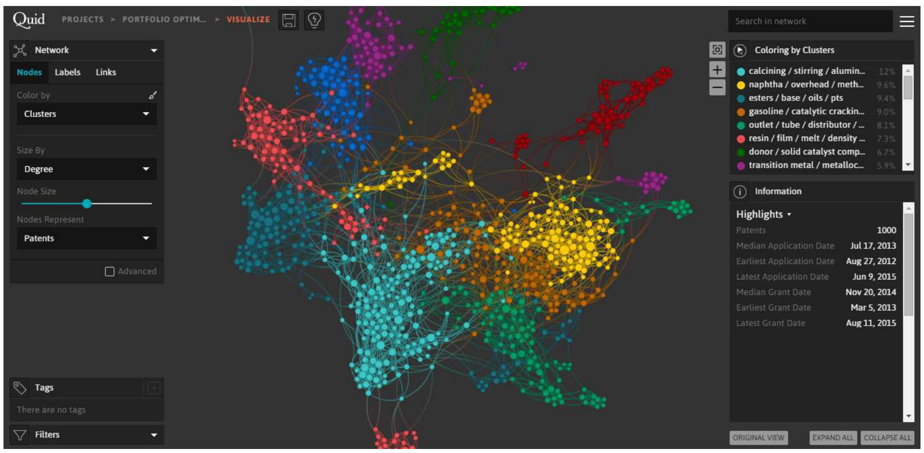
3. Proposed methodology

3.1. Patent search and retrieval

The aim of the first step of our methodology is to search and retrieve patents relevant to a given query. The Quid software [37] is used for this task. Quid is a business intelligence platform that assists decision making through visualization of complex and unstructured information. Among other features, Quid can find relevant patents according to a user query. Quid analyzes interactions among collected patents and represents patent relationships as a network map. The software uses natural language processing, text mining, network analysis and big data processing algorithms to find relations between patents and visualize related information. For instance, Figure 1 shows the patents found when searching for olefin synthesis. A node in the patent network represents a patent, while edges between nodes represent semantic similarities between patents. Two connected patents share key common language in how they describe the processes, methods, or technological solutions, signaling similar inventions. The size of a node depends on its degree, i.e., the number of connections between this node and other nodes. Larger nodes are related to a large number of other nodes and thus more representative of the nodes in their respective clusters and locales than smaller nodes. Sometimes, there are nodes called orphans that do not have any connections. Orphan nodes are nodes that are relevant to a search query and therefore included in the network, but are unique when compared to the other nodes in the network. A cluster is a group of patents that clump together since many of them are connected because they share a high degree of similar language. The density of the nodes in a cluster correlates to the average similarity between the nodes. The denser a cluster, the more similar its patents.

Quid extracts several pieces of information about each patent. The next subsection provides a detailed description of the data that we use with Quid in order to rank patents.

Figure 1: Quid's network map visualization of patents



3.2. Criteria selection

We have selected three groups of criteria in order to rank patents:

- Criteria related to a patent's connections: these criteria refer to metrics related to the number and uniqueness of a patent's connections in relationship to other patents in the same cluster or in different clusters. These criteria are:
  - Degree: the number of incident edges measures how many connections a patent has to other patents. If a patent has a high degree, it is connected to many of the patents around it, and therefore is representative of its locale. A high degree can indicate which patents are core to their respective spaces.

  - Betweenness centrality: the uniqueness of a patent's connections measures how well a patent bridges distinct parts of the network. A patent with many connections within its cluster will have a lower betweenness than a patent whose connections all reach different parts of the network. Betweenness can help us pinpoint which patents have distinct technologies (low betweenness) and which patents combine distinct technologies (high betweenness).
  - Inter-cluster fraction: the fraction of direct connections of a given patent that are within the same cluster. This fraction can be thought of as a measure of the cohesiveness of a patent's connections.

- Criteria related to a patent's neighbors: these criteria refer to metrics related to the influence and the strength of a patent's connections.
  - Flow: measures the combined strength of the patent's connections.
  - Pagerank: a proxy for a patent's influence on a network. A patent with a high pagerank is connected to many patents that, in turn, have many connections themselves.
  - Triangles: a proxy for how densely interconnected a patent is within a network. A triangle occurs when three patents are all connected to each other. A patent with a higher number of connections has the potential for a higher number of triangles.

- Criteria related to a patent's citations: these criteria refer to metrics related to the forward and backward citations of a patent.
  - Forward citations: the number of citations that a patent received from patents granted at a later time.
  - Backward citations: the number of citations given by a patent to patents granted at an earlier time

Other criteria can also be included in the proposed approach. We selected the proposed eight criteria primarily for two reasons. First, these criteria capture a considerable amount of information available about patents. Second, these criteria can be quantified with the information readily available in Quid, thus facilitating automatic retrieval information and ranking of patents. Considered separately, each criterion is likely to lead to a different ranking of the patents. The main question in this context is how to establish a mechanism for determining ways to weigh different criteria appropriately. In previous multi-criteria studies for patent ranking and evaluation, expert opinions were utilized in order to identify weights for each criterion and evaluate the patents. Such a process is subjective and likely introduces biases based on experts backgrounds. Different experts or even the same experts on different meeting times may reach different results on the same set of investigated patents. Moreover, utilization of experts does not easily lead to automation. For this reason, we propose the development of a linear optimization model that identifies weights for the different criteria in a way that intuitive constraints are satisfied. Our model is described in the next subsection.

3.3. Weight assessment using linear optimization

We address the problem of determining the relative importance of the proposed eight criteria via solution of a linear optimization model, the objective of which is to identify weights for the criteria in a way that best meets a number of constraints. For the purposes of this formulation, we begin by defining the problem variables as follows:

- w_D the weight of the criterion degree
- w_B the weight of the criterion betweenness centrality
- w_I the weight of the criterion inter-cluster fraction
- w_F the weight of the criterion flow
- w_P the weight of the criterion pagerank
- w_T the weight of the criterion triangles

- w_{FC} the weight of the criterion forward citations
- w_{BC} the weight of the criterion backward citations.

Using the above notation, we now describe the relationships/goals among the criteria weights that we would like to enforce.

- The criterion degree is very important because it indicates if a patent is representative of its locale. Additionally, the criteria related to a given patent's connections are directly related with the patent, while the criteria related to a patent's neighbors are indirectly related with the patent. Hence, the criteria related to a patent's connections are more important than the criteria related to a patent's neighbors. In other words, the sum of the weights of the criteria related to a patent's connections should be greater than or equal to the sum of the weights of the criteria related to a patent's neighbors, i.e.:

  $$w_D + w_B + w_I \geq w_F + w_P + w_T$$

- Forward and backward citations can reveal an important patent. However, we are interested in finding important newly or recently issued patents. As a result, the number of citations is not adequate to identify important recent patents since these patents have not had enough time to receive many citations. Hence, the weights of the criteria related to a patent's connections are more important than the sum of the weights of the criteria related to citations, i.e.:

  $$w_D + w_B + w_I > w_{FC} + w_{BC}$$

  In the same context, the sum of the weights of the criteria related to a patent's neighbors should be greater than or equal to the sum of the weights of the criteria related to citations, i.e.:

  $$w_F + w_P + w_T \geq w_{FC} + w_{BC}$$

Additionally, we set an upper bound of 0.25 to the sum of the weights of the criteria related to citations because we want to allow recently published patents that have not yet received citations to be highly-ranked if they describe significant innovations, i.e.:

$$w_{FC} + w_{BC} \leq 0.25$$

- Next, we compare the importance of a criterion towards the other criteria inside its group. Regarding the first group of criteria, the criterion degree is the most important one because it indicates whether a patent is representative of its locale. For this reason, we require the weight of the criterion degree to be greater than or equal to the sum of the weights of the criteria betweenness centrality and inter-cluster fraction, i.e.:

$$w_D \geq w_B + w_I$$

  Additionally, the weight of the criterion betweenness centrality is required to be less than or equal to the weight of the criterion inter-cluster fraction, i.e.:

$$w_B \leq w_I$$

- Regarding the second group of criteria, the criterion pagerank is the most important one because it shows a patent's influence on a network. For this reason, the weight of the criterion pagerank should be greater than the sum of the weights of the criteria flow and triangles, i.e.:

$$w_P > w_F + w_T$$

  Moreover, the weight of the criterion flow should be greater than the weight of the criterion triangles, i.e.:

$$w_F > w_T$$

- Regarding the third group of criteria, we consider citations received to be far more important than citations given by a patent. For this reason, we require the weight of forward citations to be at least two times greater than the weight of the backward citations, i.e.:

$$w_{FC} > 2w_{BC}$$

- Each criterion should have a nontrivial weight and no criterion should considerably outweigh other criteria. For this reason, the weight of each criterion should be greater than or equal to 0.05, i.e.:

$$w_D \geq 0.05, \quad w_B \geq 0.05, \quad w_I \geq 0.05, \quad w_F \geq 0.05,$$

$$w_P \geq 0.05, \quad w_T \geq 0.05, \quad w_{FC} \geq 0.05, \quad w_{BC} \geq 0.05$$

  and the weight of each criterion should be less than or equal to 0.33, i.e.:

$$w_D \leq 0.33, \quad w_B \leq 0.33, \quad w_I \leq 0.33, \quad w_F \leq 0.33,$$

$$w_P \leq 0.33, \quad w_T \leq 0.33, \quad w_{FC} \leq 0.33, \quad w_{BC} \leq 0.33$$

- The sum of all weights is equal to 1 (hard constraint), i.e.:

$$w_D + w_B + w_I + w_F + w_P + w_T + w_{FC} + w_{BC} = 1$$

This last constraint has been designated as a 'hard constraint,' meaning that we would like to establish weights that strictly satisfy this constraint. While desirable, all other constraints of the above formulation are 'soft,' in the sense that it would be nice to satisfy if possible. However, the formulation was developed with the understanding that it may not be possible to satisfy all of these constraints simultaneously. For this reason, the degree of satisfying soft constraints will be maximized by the objective function of the linear optimization problem. Towards this end, deviation (slack) variables are added to each soft constraint; one deviation variable is introduced in each inequality constraint and two deviation variables are introduced in each equality constraint. Additionally, the strict inequalities are modeled through the introduction of a small threshold

ϵ, that we set equal to 0.01. The final linear optimization formulation is the following:

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{25} s_i \\
\text{s.t.} \quad & w_D + w_B + w_I - w_F - w_P - w_T + s_1 \geq 0 \\
& w_D + w_B + w_I - w_{FC} - w_{BC} + s_2 \geq \epsilon \\
& w_F + w_P + w_T - w_{FC} - w_{BC} + s_3 \geq \epsilon \\
& w_{FC} + w_{BC} + s_4 \leq 0.25 \\
& w_D - w_B - w_I + s_5 \geq 0 \\
& w_B - w_I + s_6 \geq 0 \\
& w_P - w_F - w_T + s_7 \geq \epsilon \\
& w_F - w_T + s_8 \geq \epsilon \\
& w_{FC} - 2w_{BC} + s_9 \geq 0 \\
& w_D + s_{10} \geq 0.05 \\
& w_B + s_{11} \geq 0.05 \\
& w_I + s_{12} \geq 0.05 \\
& w_F + s_{13} \geq 0.05 \\
& w_P + s_{14} \geq 0.05 \\
& w_T + s_{15} \geq 0.05 \\
& w_{FC} + s_{16} \geq 0.05 \\
& w_{BC} + s_{17} \geq 0.05 \\
& w_D + s_{18} \leq 0.33 \\
& w_B + s_{19} \leq 0.33 \\
& w_I + s_{20} \leq 0.33 \\
& w_F + s_{21} \leq 0.33 \\
& w_P + s_{22} \leq 0.33 \\
& w_T + s_{23} \leq 0.33
\end{aligned}
$$

$$w_{FC} + s_{24} \leq 0.33$$
$$w_{BC} + s_{25} \leq 0.33$$
$$w_D + w_B + w_I + w_F + w_P + w_T + w_{FC} + w_{BC} = 1$$
$$w_D, w_B, w_I, w_F, w_P, w_T, w_{FC}, w_{BC} \geq 0$$
$$s_j \geq 0, \quad j = 1, \ldots, 25$$

This linear optimization problem was solved with the optimization solver CPLEX v12.6.3. The value of the objective function (the total convergence deviation) was equal to 0, meaning that all constraints were fully satisfied. Interestingly, alternative optimal solutions were identified. Two indicative solutions were selected for experimentation and are shown in Table 2. They will be referred to as sets of weights A and B.

While subjective, the above constraints and goals are intuitive. Additionally, decision makers can modify the goals based on their beliefs about relative priorities among the criteria. Customized linear or nonlinear constraints can be included if more relationships among the criteria are discovered. The observed relevance of the eight criteria is influenced to a certain degree by the applied optimization scheme for their weights. In Section 5.6, we validate the use of eight criteria for patent ranking in a comparison between configurations with a varying number of criteria and compare the rankings obtained by the different approaches. It will be seen that configurations with fewer than five criteria can find only a subset of the most important patents.

Equipped with the weights of the different criteria, we can proceed to rank patents. A naive approach in this context would be to rank each patent according to the weighted sum of the eight criteria. A downside of this approach is that it ignores the different scales of the criteria involved [38]. Even if we use a normalization technique to eliminate the units of the criteria, a poor value for one criterion can be heavily outweighed by a very good value for another criterion. Additionally, using an additive aggregation operator, such as weighted sum, is equivalent to assuming that all the criteria are independent [39]. How-

Table 2: Criteria weights

Variable	Criterion	Weight	
		Solution A	Solution B
w_D	Degree	0.33	0.33
w_B	Betweenness centrality	0.05	0.05
w_I	Inter-cluster fraction	0.05	0.05
w_F	Flow	0.08	0.06
w_P	Pagerank	0.17	0.31
w_T	Triangles	0.07	0.05
w_{FC}	Forward citations	0.17	0.10
w_{BC}	Backward citations	0.08	0.05

ever, in practice this is usually not feasible. To address these issues, in the next subsection, we utilize the weights of Table 2 in the multicriteria optimization procedure TOPSIS [23].

3.4. TOPSIS method

TOPSIS [40, 23] stands for Technique of Order Preference Similarity to the Ideal Solution. This method has been successfully applied in multicriteria decision making in many application areas, including supply chain management, logistics, chemical engineering, and patent rating [41]. In the latter case, experts provided the weights and different criteria were used compared to the ones proposed here.

For each candidate solution of a multicriteria optimization problem, the TOPSIS method calculates the distances from an ideal and anti-ideal solution. The ideal solution is defined as the one that optimizes all criteria simultaneously, while the anti-ideal solution corresponds to a worst-case point for all criteria. The primary purpose of TOPSIS is to rank solutions based on their relative closeness to the ideal solution. First, a scaling procedure is used to account for the different scales of different criteria.

Let us assume that a multiple criteria decision making problem has m alternatives, $A_1, \ldots, A_m$, and n decision criteria, $C_1, \ldots, C_n$. In our application, m is the number of patents to be ranked, while n corresponds to the number of criteria proposed above, i.e., $n = 8$. First, each alternative is evaluated according to each of the n criteria. These evaluations form a decision matrix $X = (x_{ij})_{m \times n}$. Similar to the previous subsection, let $W = (w_1, \ldots, w_n)$ be the vector of the criteria weights, where $\sum_{j=1}^{n} w_j = 1$.

The TOPSIS method involves the following five steps:

- **Step 1. Calculation of the weighted normalized decision matrix**. The first step is to normalize the decision matrix in order to eliminate the effect of different criteria units. The normalized decision matrix is computed using the following vector normalization technique:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}}, \quad i = 1, \ldots, m; \quad j = 1, \ldots, n$$

  Then, the normalized decision matrix is multiplied by the weight associated with each of the criteria. The normalized weighted decision matrix is calculated as follows:

$$v_{ij} = w_j r_{ij}, \quad i = 1, \ldots, m; \quad j = 1, \ldots, n$$

- **Step 2. Determination of the ideal and anti-ideal solutions**. The ideal (A^+) and anti-ideal (A^-) solutions are computed as follows:

$$A^+ = \left(v_1^+, \ldots, v_n^+\right); \quad v_i^+ = \max_j v_{ij}, \quad i = 1, \ldots, m$$

$$A^- = \left(v_1^-, \ldots, v_n^-\right); \quad v_i^- = \min_j v_{ij}, \quad i = 1, \ldots, m$$

- **Step 3. Calculation of the distance from the ideal and anti-ideal solutions**. The distance from the ideal and the anti-ideal solution is computed for each alternative as follows:

$$D_i^+ = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^+\right)^2}, \quad i = 1, \ldots, m$$

$$D_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^-\right)^2}, \quad i = 1, \ldots, m$$

- **Step 4. Calculation of the relative closeness to the ideal solution**. The relative closeness of each alternative to the ideal solution is calculated as follows:
$$C_i = \frac{D_i^-}{D_i^+ + D_i^-}, \quad i = 1, \ldots, m$$
where $0 \leq C_i \leq 1$.

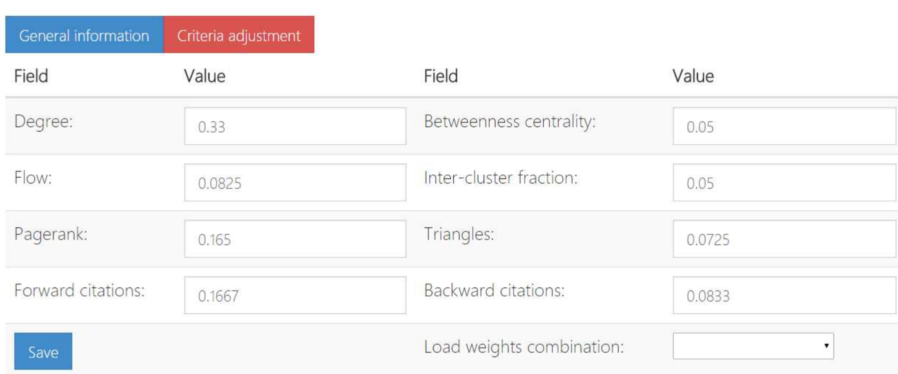
- **Step 5. Ranking the alternatives**. The alternatives are ranked from best (highest relative closeness value C_i) to worst.

4. Decision support system

In order to facilitate application and validation of the proposed methodology, we have implemented a web-based decision support system (DSS). The DSS has been implemented using PHP, MySQL, Ajax, and jQuery. Figure 2 presents the decision-making process that a decision maker can follow in order to retrieve and rank patents according to a specific query. Initially, the decision maker submits a query to find relevant patents using Quid. Then, the decision maker reviews the collected patents and excludes those that appear irrelevant. In the next step, the decision maker exports a json file that contains the identified information about the patents. This information becomes the input in our decision support system. The decision maker can use the default weights of the criteria obtained above (Table 2) or specify different values (see Figure 3). Then, the system extracts the available information for each patent and ranks the patents according to the TOPSIS method. Finally, the rankings are presented to the decision maker (Figure 4). The decision maker can choose to display the top-ranked patents or get a thorough report of the rankings. Detailed results are also displayed for each patent (Figure 5). Additionally, the following information is available for each patent: (i) index number, (ii) title, (iii) status, (iv) abstract, (v) link, and (vi) the values for each of the eight criteria presented in Section 3.2.

Figure 2: Decision making process

Figure 3: Adjusting weights of the criteria



General information | Criteria adjustment

Field	Value	Field	Value
Degree:	0.33	Betweenness centrality:	0.05
Flow:	0.0825	Inter-cluster fraction:	0.05
Pagerank:	0.165	Triangles:	0.0725
Forward citations:	0.1667	Backward citations:	0.0833
Save		Load weights combination:	

Figure 4: Patent ranking

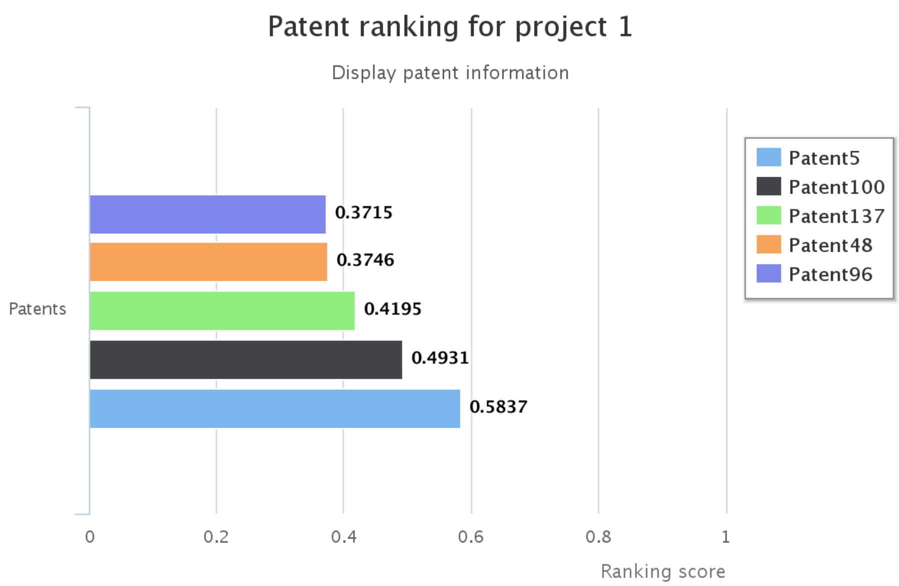
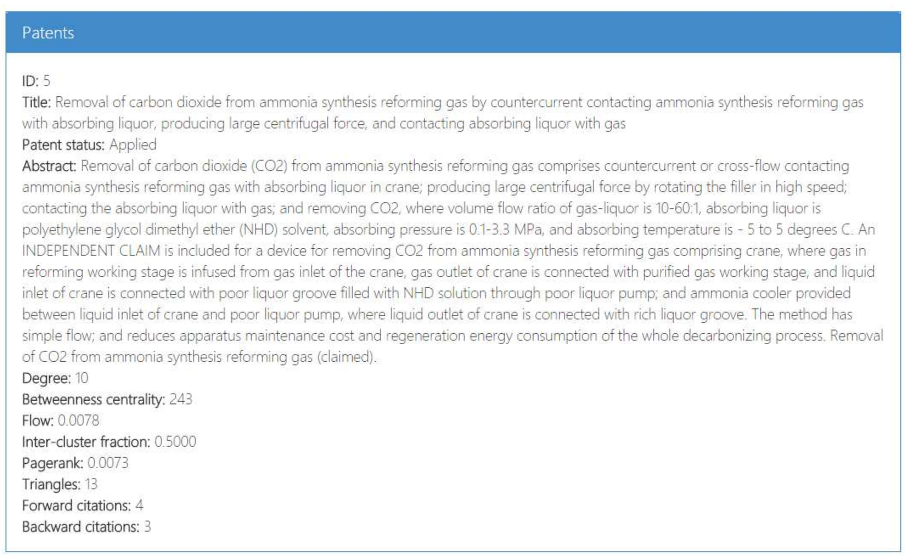


Figure 5: Patent information


Patents

ID: 5
Title: Removal of carbon dioxide from ammonia synthesis reforming gas by countercurrent contacting ammonia synthesis reforming gas with absorbing liquor, producing large centrifugal force, and contacting absorbing liquor with gas
Patent status: Applied
Abstract: Removal of carbon dioxide (CO2) from ammonia synthesis reforming gas comprises countercurrent or cross-flow contacting ammonia synthesis reforming gas with absorbing liquor in crane; producing large centrifugal force by rotating the filler in high speed; contacting the absorbing liquor with gas; and removing CO2, where volume flow ratio of gas-liquor is 10-60:1, absorbing liquor is polyethylene glycol dimethyl ether (NHD) solvent, absorbing pressure is 0.1-3.3 MPa, and absorbing temperature is - 5 to 5 degrees C. An INDEPENDENT CLAIM is included for a device for removing CO2 from ammonia synthesis reforming gas comprising crane, where gas in reforming working stage is infused from gas inlet of the crane, gas outlet of crane is connected with purified gas working stage, and liquid inlet of crane is connected with poor liquor groove filled with NHD solution through poor liquor pump; and ammonia cooler provided between liquid inlet of crane and poor liquor pump, where liquid outlet of crane is connected with rich liquor groove. The method has simple flow; and reduces apparatus maintenance cost and regeneration energy consumption of the whole decarbonizing process. Removal of CO2 from ammonia synthesis reforming gas (claimed).
Degree: 10
Betweenness centrality: 243
Flow: 0.0078
Inter-cluster fraction: 0.5000
Pagerank: 0.0073
Triangles: 13
Forward citations: 4
Backward citations: 3

In the next section, we describe the results from using this DSS in the context of an application area that motivated this research.

5. Computational experiments

We illustrate the proposed methodology to find the most novel patents that describe new processes to produce three well-known chemicals:

1. ammonia process synthesis
2. olefin synthesis
3. polyethylene synthesis.

After exporting patent data from Quid, we upload each case to our DSS. Three different sets of weights, denoted here as sets A, B and C, were used to rank the patents for each case. Sets A and B correspond to the two solution points of the linear optimization formulation presented in Section 3.4. Set C involves equal weights ($w = 0.125$ for all criteria weights). The following subsections present the results for each of the three cases. Subsequently, we also investigate the relationships between the criteria utilized in this paper.

5.1. Ammonia process synthesis

Initially, we search for patents that describe new processes of producing ammonia. As a very important and fundamental industrial chemical product, ammonia is widely used in both its pure form and as a feedstock for a wide variety of other chemical products like synthetic polymers, nitric acid, and commercial fertilizers. Today, many of the chemical plants produce ammonia using the traditional Haber process where hydrogen and nitrogen are combined directly under high pressure and high temperature, which is a high energy consumption process. The word 'process' is defined generally, which means that we are looking for patents with a wide range of subjects that not only include new reactions but also new catalysts, new apparatuses or even new control schemes. Hence, the keywords that we defined in Quid are: ammonia, ammonia production, manufacture of ammonia, produce ammonia, ammonia process, and ammonia synthesis. A total of 166 patents that are closely related to our objective were exported from Quid.

Table 3: Top five patents in ammonia process synthesis

Set of weights	Top five patents				
	First	Second	Third	Fourth	Fifth
A	5	100	137	48	96
B	5	100	137	48	96
C	137	5	20	100	48

The top five ranked patents for all sets of weights are presented in Table 3. For instance, the top five patents using weights set A are Patents 5, 100, 137, 48, and 96. The rankings using weights A and B are the same. The ranking changes if the third set of weights (set C) is used. However, Patents 137, 5, 100, and 48 are still in the top five. Only Patent 96 drops out of the top five. This means that most high-ranked patents of this example are very important and their ranking does not change drastically even if we use different sets of weights. Among the high-ranked patents, Patents 5, 100, and 137 deal with generating syngas used in ammonia processes, covering topics from syngas purification to new methods for producing syngas. These observations are consistent with common knowledge that syngas plays an essential role in ammonia processes. Thus, the rankings of patents in this area provide high-quality guidance for decision makers interested in ammonia process design.

5.2. Olefin synthesis

The second example deals with the synthesis of olefins, which form important intermediates in the production of many chemicals. Olefins represent a crucial group of petrochemicals. They are mostly produced by steam cracking of petroleum fractions. Olefins include ethylene, propylene, and butadiene. Because of olefins' wide range of uses and because they are produced from petroleum, the keywords that we used in Quid were: olefins production, produce olefin, and olefin process synthesis. A total of 102 patents that are closely related to our objective were exported from Quid.

Table 4: Top five patents in olefin synthesis

Set of weights	Top five patents				
	First	Second	Third	Fourth	Fifth
A	75	76	22	45	55
B	75	22	76	45	17
C	75	45	22	17	76

The five top-ranked patents for all weight sets are presented in Table 4. The patents included in the top five are almost identical with all sets of weights. However, their ranking is different. For instance, Patent 22 is ranked higher if weights set B is used. Additionally, although Patent 75 does not have the highest values on all criteria, it is always ranked first on all sets of weights. This is because TOPSIS initially normalizes the decision matrix and then ranks the alternatives based on their normalized data. Patent 75 claims a new catalyst for producing an olefin-based polymer. As an intermediate chemical product, olefins are the basis for polymers used in plastics, resins, fibers, lubricants, and gels. Moreover, in the olefin industry, many processes are barely profitable without using catalysts. Hence, the high ranking of this patent suggests that catalyst development will be a crucial trend for the chemical industry in the future. Among the top five patents, Patents 17 and 45 are also related with catalyst composition, further validating the quality of the proposed methodology.

5.3. Polyethylene synthesis

The last example deals with polyethylene synthesis. Polyethylene is one of the most common plastics in today's world. It is primarily used in packaging (plastic bags, plastic films, bottles, etc.) because of its good chemical resistance. Based on its density and branching, polyethylene can be classified in three major categories: linear low-density polyethylene (LLDPE), low-density polyethylene (LDPE), and high-density polyethylene (HDPE). Like many other commercial polymers, polyethylene can be produced by addition polymerization of ethylene, i.e., building long molecular chains comprised of ethylene monomers. The keywords that we defined in Quid were: polyethylene production, ethylene polymerization, LDPE, and HDPE. A total of 116 patents that are closely related to our objective were exported from Quid.

The five top-ranked patents for all sets of weights are presented in Table 5. The patents included in the top four are identical for the first two sets of weights. When using equal weights, the ranking changes but four of the top five patents with weights A are still in the top five. The only exception is Patent 37, which has been replaced by Patent 67. Three patents persist across all three weights, suggesting the importance of these patents. Patents 32 and 36 describe processes that produce high molecular weight polyethylene. As mentioned before, high-density polyethylene is one of the major polyethylene classes. This particular type of polyethylene is of great commercial importance and represents an essential product in polyethylene production. Patent 38 does not deal with polymerization but is related with the production of polyethylene powder. As the final production steps, drying and pelletizing determine the quality of the final product. Furthermore, we notice that Patent 38 is ranked fourth on all weight sets. Therefore, we can conclude that the pelletizing step is almost as important as the polymerization step. When we rank the patents with equal weights, the patent that is ranked first by weights A and B is replaced by Patent 93, which introduces a method that is useful for gas-phase polymerization. Although gas-phase polymerization is very common, there are still limits for this technology because of the wide variety of polyethylene products that are produced. By selecting this particular patent as the top patent, the rankings with the equal weights set might mislead the decision makers in this case.

5.4. Relationships between ranking criteria

The results for the three chemicals reveal that similar rankings are produced by the two optimal sets of weights in Table 2. Moreover, the high-ranked patents were found to be those that represent the core processes in their respective field. In this subsection, we investigate potential relationships between the criteria

Table 5: Top five patents in polyethylene synthesis

Set of weights	Top five patents				
	First	Second	Third	Fourth	Fifth
A	36	32	37	38	93
B	36	32	37	38	33
C	93	36	67	38	32

utilized in this paper. We introduce perturbations to the weights of a small set of criteria and study their impact on the final ranking of patents. The first set of weights (A) is used as the basis in the ammonia synthesis, the second set (B) in the olefin synthesis, and the third set (C) is used in the polyethylene synthesis.

5.4.1. Degree versus flow

The first set of criteria that will be analyzed are the degree and the flow. Although these criteria are not categorized in the same group, they both measure the connections and patent relationships to a certain degree. The degree counts the number of patent connections and the flow measures the combined strength of these connections. Table 6 presents a comparison of different scenarios if we introduce perturbations to the weights of these two criteria while the sum of their weights remains the same. The second and third row of Table 6 include the pairs of weights used to investigate the relationship between degree and flow, while the other rows report the patents for each scenario. For the most part, the rankings stay unchanged. This confirms our hypothesis that these two criteria function similarly in ranking. Hence, changing the weights of these two criteria should have little effect on the patent rankings.

5.4.2. Inter-cluster fraction versus betweenness centrality

While degree and flow provide a relatively similar measurement of a patent's connections, the inter-cluster fraction and betweenness centrality refer to two opposite metrics. Inter-cluster fraction is a ratio of how many direct connec-

Table 6: Relationship between degree and flow

Ranking \ Criteria	Case study	Ammonia synthesis			Olefin synthesis			Polyethylene synthesis		
	Degree	0.3300	0.4100	0.2300	0.33	0.38	0.23	0.125	0.225	0.025
	Flow	0.0825	0.0025	0.1825	0.06	0.01	0.16	0.125	0.025	0.225
1		5	5	5	75	75	75	93	93	93
2		100	100	100	22	22	22	36	36	36
3		137	137	137	76	76	76	67	67	67
4		48	48	48	45	45	17	38	38	38
5		96	96	96	17	17	45	32	32	34

tions of a given patent are present within the same cluster of the patent, while betweenness centrality measures the uniqueness of a patent's connections. A patent with many connections within its cluster will have a lower betweenness in comparison to a patent whose connections reach different parts of the network. Hence, these two criteria drive patent rankings in opposite ways. The higher the inter-cluster fraction, the lower the betweenness centrality. Table 7 presents a comparison of different scenarios if we introduce perturbations to the weights of these two criteria while the sum of their weights remains the same. Although these two criteria act in opposite ways, the rankings are not sensitive to the changes of their weights. For example, when we change the weights of inter-cluster fraction and betweenness centrality in the ammonia process synthesis case, the top five patents remain the same. This happens because other criteria are likely to dominate the rankings, since the sum of the two criteria weights is only 10%. In the polyethylene synthesis example, the rankings are quite different after changing the weight of betweenness centrality from 0.125 to 0.225 while decreasing the weight of inter-cluster fraction, thus suggesting that there are cases in which these criteria are important to account for.

5.4.3. Forward citations versus backward citations

The last set of criteria compared involves the forward and backward citations. Citation counts are key measurements that reflect a patent's value. Unlike other criteria, forward and backward citations are strictly independent. Therefore, if we perform perturbations on their weights, we expect the resulting rankings to be random and unpredictable. Table 8 presents a comparison of different scenarios when we introduce perturbations to the weights of these two criteria while keeping their sum constant. The table shows that the rankings and even the identities of the five top-ranked patents change as the weights change. The results confirm our expectation that there is no relationship between the two criteria.

Table 7: Relationship between fraction and betweenness centrality

	Case study	Ammonia synthesis			Olefin synthesis			Polyethylene synthesis		
Criteria	**Inter-cluster fraction**	0.05	0.09	0.01	0.05	0.09	0.01	0.125	0.225	0.025
Ranking	**Betweenness centrality**	0.05	0.01	0.09	0.05	0.01	0.09	0.125	0.025	0.225
1		5	5	5	75	75	75	93	93	67
2		100	100	100	22	22	22	36	36	96
3		137	137	137	76	76	76	67	38	93
4		48	48	48	45	45	17	38	34	24
5		96	96	96	17	8	45	32	37	36

Table 8: Relationship between forward and backward citations

	Case study	Ammonia synthesis			Olefin synthesis			Polyethylene synthesis		
Criteria	**Forward citations**	0.1667	0.2467	0.0667	0.10	0.14	0.01	0.125	0.225	0.025
Ranking	**Backward citations**	0.0833	0.0033	0.1833	0.05	0.01	0.14	0.125	0.025	0.225
1		5	5	137	75	75	45	93	36	93
2		100	100	20	22	76	17	36	32	38
3		137	48	5	76	22	22	67	37	34
4		48	96	100	45	55	7	38	33	67
5		96	134	92	17	8	3	32	38	58

Table 9: Top five patents using the weighted sum method

Case study \ Set of weights	Top five patents				
	First	Second	Third	Fourth	Fifth
Ammonia process synthesis \ A	150	66	69	142	146
Olefin synthesis \ B	56	51	20	71	17
Polyethylene synthesis \C	67	96	24	68	85

5.5. Comparisons between naive approaches and multiple criteria methods

In the three examples above, we compared the use of the proposed weights of Table 2 against the use of equal weights for all the proposed criteria. Although the naive approach of using equal weights identified many of the top-ranked patents, it sometimes missed important patents. This observation justifies the use of the proposed linear optimization model and corresponding optimal weights for the ranking of patents. Another naive approach to multiple criteria decision making would be to use any set of weights but not in conjunction with TOPSIS. TOPSIS would scale the weights and rank patents based on distances from ideal and anti-ideal patents. In Table 9, we show the results of this naive approach in the context of the three cases. The results of this table were obtained with each of the three sets of weights used in the three case studies above to rank patents. In all cases, we used the weighted sum of the eight criteria as the sole comparison metric.

In the ammonia process synthesis example, all top-ranked patents are different than those obtained using TOPSIS. This unique ranking was generated because all top-ranked patents using the weighted sum method have very large values in the betweenness centrality criterion and low values for other criteria. For instance, all top-ranked patents have zero forward citations. Although the weight for betweenness centrality is only 0.05, its weighted contribution to the comparison metric is still very large and influences the ranking. Thus, the ranking is of poor quality. For instance, Patent 150 claims a catalyst performance testing device used in ammonia process synthesis. Producing ammonia is a profitable industrial process that requires catalysis. One important issue during this process is the catalyst's deactivation over the time. In order to maintain high effectiveness and efficiency, we need to replace the catalyst periodically. The catalyst performance testing device described in this patent is practical and useful for all ammonia processes. However, many similar technologies exist and their improvements are no longer considered breakthrough developments. Even though this patent is of some importance, it should not be ranked first. This patent is ranked 24th when TOPSIS is used.

The situation is similar in the olefin synthesis case study since the betweenness centrality values of the top-ranked patents are also very large. Additionally, some patents introduce almost opposite chemical reaction directions that may confuse and mislead the decision maker. For instance, Patent 56 claims an olefin production system to produce olefins where the olefins are used to produce alkanes later. On the other hand, Patent 20 introduces a catalytic cracking scheme that uses a lower alkane to produce olefins.

In the polyethylene synthesis case study, the performance of the weighted sum method does not improve even if we use equal weights. This is because the betweenness centrality values of the top-ranked patents are still very large. After multiplying the criterion value with the corresponding weight (0.125), the resulting product accounts for over 95 percent of the weighted summation value. The ranking result is of poor quality. For instance, Patent 85, which claims a high-density polyethylene polymeric composition for producing containers is ranked fifth using the weighted sum method. This patent is ranked 24th when TOPSIS is used. Considering that this patent claims a polymeric composition for specific use, its high ranking may have little value to decision makers who consider industrial pipe applications.

As seen in this section, if we use the weighted sum method to rank the patents, we obtain different rankings because they are largely biased by the criteria that have very large values and dominate all others. On the other hand, TOPSIS measures the distance of each patent from the ideal solution, thus providing us with patents close to the ideal solution. Another interesting finding is that all high-ranked patents using the weighted sum method have only a few forward citations. With citations as an evaluation metric, it is hard to believe that the patents with a small number of citations are promising. In the ammonia process synthesis example, Patent 5 (top-ranked using TOPSIS), dealing with the purification of syngas, should be more promising than Patent 150 (top-ranked using the weighted sum method) that claims a catalyst performance testing device. By purifying syngas, one can shorten the catalyst replacement times and reduce operating costs. In the polyethylene synthesis case, Patent 36 (top-ranked using TOPSIS) has 11 forward citations and seven backward citations. This patent claims a process that produces high molecular weight polyethylene. For the same case study, Patent 67 (top-ranked using the weighted sum method) has one forward citation and 18 backward citations. This granted patent claims a process that is not only useful for polyethylene but also for ethylene copolymer, which is not related to the subject of the study. These observations justify the use of TOPSIS for ranking patents.

5.6. Parametric analysis of the effect of the number of criteria

In the three case studies above, we used eight criteria to evaluate and rank patents. In this subsection, we address the question whether similar quality rankings can be obtained by fewer than eight metrics. We use twenty two design projects in chemical R&D analytics that have been used in capstone design courses [42, 43]. For each case, we investigate the quality of rankings by varying the number of criteria ($1, \cdots, 7$ criteria) and comparing the rankings (top five and top ten) with the rankings obtained by the proposed method (eight criteria using equal weights).

Table 10 presents the results from this comparison. We calculate the same top 5 and top 10 patents found by each of the configurations as a percentage of the top 5 / 10 found by the proposed configuration. In the configurations with multiple criteria ($2, \cdots, 7$), we perform all criteria combinations and report the one that has the most similar patents with the proposed configuration. The reported best configurations are the following:

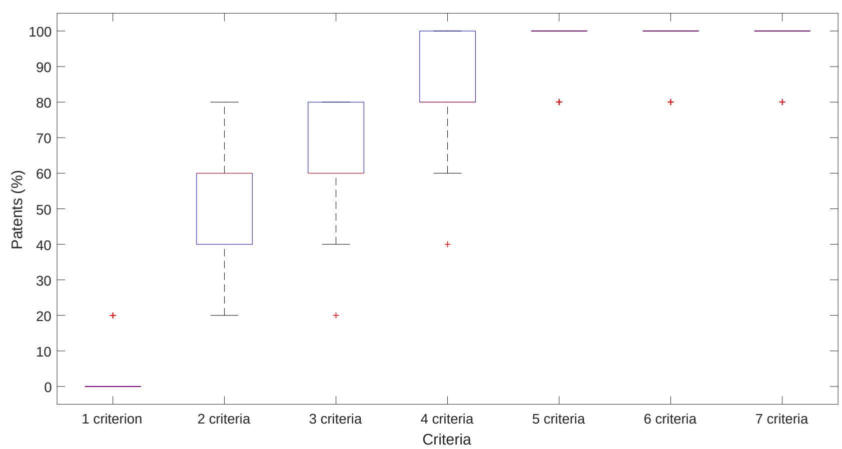
1. The best single-criterion configuration uses pagerank.
2. The best configuration with two criteria uses pagerank and forward citations.
3. The best configuration with three criteria uses pagerank, forward citations, and degree.
4. The best configuration with four criteria uses pagerank, forward citations, degree, and backward citations.
5. The best configuration with five criteria uses pagerank, forward citations, degree, backward citations, and betweeness centrality.
6. The best configuration with six criteria uses pagerank, forward citations, degree, backward citations, betweeness centrality, and triangles.
7. The best configuration with seven criteria uses pagerank, forward citations, degree, backward citations, betweeness centrality, triangles, and flow.

Figures 6 and 7 depict graphically the degree dispersion and skewness in the results presented in Table 10. All these results reveal that the configurations with fewer than five criteria cannot find all of the important patents (at least 15% and 25% of the top 5 and top 10 patents, respectively, cannot be found), while the configurations with more than or equal to five criteria can find almost all of the important patents. Hence, we can identify mostly the same set of top patents using at least the five most important criteria, i.e., pagerank, forward citations, degree, backward citations, and betweenness centrality. The inclusion of the other three criteria yields to the identification of 5% and 8% more of the top 5 and top 10 patents, respectively. When considering that the proposed methodology with eight criteria only needs an average of 0.6 seconds to find the most important patents (compared to 0.5 seconds for the configuration with five criteria), it is worth using eight criteria on the proposed multicriteria method.

Table 10: Effect of the number of criteria on the number of top patents identified

Number of criteria	Top five			Top ten		
	Maximum	Minimum	Average	Maximum	Minimum	Average
1	20%	0%	3%	40%	0%	10%
2	80%	20%	53%	70%	30%	52%
3	80%	20%	63%	80%	30%	65%
4	100%	40%	85%	90%	40%	75%
5	100%	80%	95%	100%	80%	92%
6	100%	80%	96%	100%	90%	96%
7	100%	80%	97%	100%	90%	96%

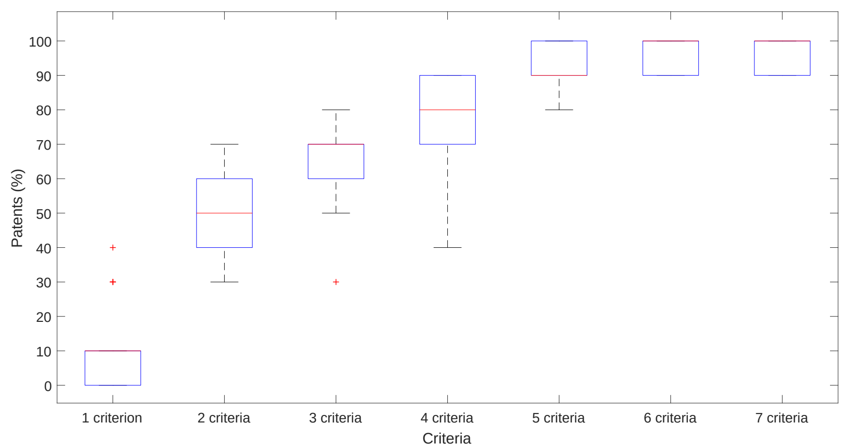
Figure 6: Patents on the top five by each configuration



6. Conclusions

We presented a methodology to rank patents based on multiple criteria and an intuitive linear optimization formulation that reveals how to weigh different criteria. We also implemented a web-based decision support system to automate the proposed methodology. This system was used to validate the methodology in the context of finding the most important patents that describe new processes for three well-known chemicals: (i) ammonia process synthesis, (ii) olefin synthesis,

Figure 7: Patents on the top ten by each configuration



and (iii) direct propylene oxidation. The proposed methodology identified novel patents for all three case studies. The high-ranked patents that were identified are those that represent the core processes in their respective field. A sensitivity analysis revealed that the rankings are relatively stable when either of the two proposed optimal sets of weights is used. Finally, we validated the use of eight criteria to rank patents in a comparison between configurations with a varying number of criteria and comparing the rankings with the rankings obtained by the proposed method. The configurations with fewer than five criteria can find only a subset of the most important patents. Using five criteria made it possible to identify 95% of the top patents identified when using all eight proposed criteria. These results suggest that multiple criteria should be utilized in patent ranking.

References

[1] R. S. Campbell, Patent trends as a technological forecasting tool, World Patent Information 5 (3) (1983) 137–143.

[2] L. Zhang, L. Li, T. Li, Patent mining: A survey, ACM SIGKDD Explorations Newsletter 16 (2) (2015) 1–19.

[3] S. Sheremetyeva, Natural language analysis of patent claims, in: Proceedings of the ACL-2003 workshop on Patent corpus processing-Volume 20, Association for Computational Linguistics, 2003, pp. 66–73.

[4] A. Shinmori, M. Okumura, Y. Marukawa, M. Iwayama, Patent claim processing for readability: Structure analysis and term explanation, in: Proceedings of the ACL-2003 workshop on Patent corpus processing-Volume 20, Association for Computational Linguistics, 2003, pp. 56–65.

[5] B. Stein, D. Hoppe, T. Gollub, The impact of spelling errors on patent search, in: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 2012, pp. 570–579.

[6] K. Konishi, Query terms extraction from patent document for invalidity search, in: Proc. of NTCIR, Vol. 5, 2005.

[7] P. Mahdabi, M. Keikha, S. Gerani, M. Landoni, F. Crestani, Building queries for prior-art search, in: A. Hanbury, A. Rauber, A. de Vries (Eds.), Multidisciplinary Information Retrieval, Vol. 6653 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2011, pp. 3–15.

[8] W. Tannebaum, A. Rauber, Analyzing query logs of USPTO examiners to identify useful query terms in patent documents for query expansion in patent searching: A preliminary study, in: Multidisciplinary Information Retrieval, Springer, 2012, pp. 127–136.

[9] J. Xu, W. B. Croft, Query expansion using local and global document analysis, in: Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, 1996, pp. 4–11.

[10] L. Zhang, Z. Liu, L. Li, C. Shen, T. Li, Patsearch: an integrated framework for patentability retrieval, Knowledge and Information Systems (2017) 1–24.

[11] M.-H. Huang, L.-Y. Chiang, D.-Z. Chen, Constructing a patent citation map using bibliographic coupling: A study of Taiwan's high-tech companies, Scientometrics 58 (3) (2003) 489–506.

[12] J. Tang, B. Wang, Y. Yang, P. Hu, Y. Zhao, X. Yan, B. Gao, M. Huang, P. Xu, W. Li, et al., PatentMiner: Topic-driven patent analysis and mining, in: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2012, pp. 1366–1374.

[13] B. Yoon, Y. Park, A text-mining-based patent network: Analytical tool for high-technology trend, The Journal of High Technology Management Research 15 (1) (2004) 37–50.

[14] B.-U. Yoon, C.-B. Yoon, Y.-T. Park, On the development and application of a self–organizing feature map–based patent map, R&D Management 32 (4) (2002) 291–300.

[15] J. Yoon, W. Seo, B.-Y. Coh, I. Song, J.-M. Lee, Identifying product opportunities using collaborative filtering-based patent analysis, Computers & Industrial Engineering 107 (2017) 376–387.

[16] S. Lee, B. Yoon, Y. Park, An approach to discovering new technology opportunities: Keyword-based patent map approach, Technovation 29 (6) (2009) 481–497.

[17] Y. G. Kim, J. H. Suh, S. C. Park, Visualization of patent analysis for emerging technology, Expert Systems with Applications 34 (3) (2008) 1804–1812.

[18] A. Fujii, Enhancing patent retrieval by citation analysis, in: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, 2007, pp. 793–794.

[19] P. Hu, M. Huang, P. Xu, W. Li, A. K. Usadi, X. Zhu, Finding nuggets in IP portfolios: Core patent mining through textual temporal analysis, in:

Proceedings of the 21st ACM international conference on Information and knowledge management, ACM, 2012, pp. 1819–1823.

[20] Y. Liu, P.-Y. Hseuh, R. Lawrence, S. Meliksetian, C. Perlich, A. Veen, Latent graphical models for quantifying and predicting patent quality, in: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2011, pp. 1145–1153.

[21] M. B. Albert, D. Avery, F. Narin, P. McAllister, Direct validation of citation counts as indicators of industrially important patents, Research Policy 20 (3) (1991) 251–259.

[22] P. Ellis, G. Hepburn, C. Oppenhein, Studies on patent citation networks, Journal of Documentation 34 (1) (1978) 12–20.

[23] C.-L. Hwang, K. Yoon, Multiple attributes decision making methods and applications, Springer-Verlag, Berlin, 1981.

[24] M. Trajtenberg, A penny for your quotes: Patent citations and the value of innovations, The RAND Journal of Economics (1990) 172–187.

[25] D. Harhoff, F. Narin, F. M. Scherer, K. Vopel, Citation frequency and the value of patented inventions, Review of Economics and Statistics 81 (3) (1999) 511–515.

[26] B. H. Hall, A. B. Jaffe, M. Trajtenberg, The NBER patent citation data file: Lessons, insights and methodological tools, Tech. rep., National Bureau of Economic Research (2001).

[27] B. H. Hall, A. Jaffe, M. Trajtenberg, Market value and patent citations, The RAND Journal of Economics (2005) 16–38.

[28] M. A. Hasan, W. S. Spangler, T. Griffin, A. Alba, Coa: Finding novel patents through text analysis, in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2009, pp. 1175–1184.

[29] X. Jin, S. Spangler, Y. Chen, K. Cai, R. Ma, L. Zhang, X. Wu, J. Han, Patent maintenance recommendation with patent information network model, in: Data Mining (ICDM), 2011 IEEE 11th International Conference on, IEEE, 2011, pp. 280–289.

[30] S. Oh, Z. Lei, P. Mitra, J. Yen, Evaluating and ranking patents using weighted citations, in: Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries, ACM, 2012, pp. 281–284.

[31] A. Barbazza, M. Collan, M. Fedrizzi, P. Luukka, Consensus modeling in multiple criteria multi-expert real options-based valuation of patents, in: Intelligent Systems 2014, Springer, 2015, pp. 269–278.

[32] M. Collan, M. Fedrizzi, P. Luukka, A multi-expert system for ranking patents: An approach based on fuzzy pay-off distributions and a TOPSIS–AHP framework, Expert Systems with Applications 40 (12) (2013) 4749–4759.

[33] Y. Lawryshyn, M. Collan, P. Luukka, M. Fedrizzi, Valuing patents utilizing consensual dynamics in a real options framework, in: IIE Annual Conference. Proceedings, Institute of Industrial Engineers-Publisher, 2014, p. 486.

[34] S. Oh, Z. Lei, W.-C. Lee, J. Yen, Patent evaluation based on technological trajectory revealed in relevant prior patents, in: Advances in Knowledge Discovery and Data Mining, Springer, 2014, pp. 545–556.

[35] B. Wang, C.-H. Hsieh, Measuring the value of patents with fuzzy multiple criteria decision making: Insight into the practices of the Industrial Technology Research Institute, Technological Forecasting and Social Change 92 (2015) 263–275.

[36] Y. Zhang, Y. Qian, Y. Huang, Y. Guo, G. Zhang, J. Lu, An entropy-based indicator system for measuring the potential of patents in technological innovation: rejecting moderation, Scientometrics 111 (3) (2017) 1925–1946.

[37] Quid Inc., Business Intelligence Software, http://quid.com, accessed: 01-09-2018.

[38] E. Triantaphyllou, Multi-criteria decision making methods: A comparative study, Vol. 44, Springer Science & Business Media, New York, 2013.

[39] T. Magoč, F. Modave, The optimality of non-additive approaches for portfolio selection, Expert Systems with Applications 38 (10) (2011) 12967–12973.

[40] K. Yoon, Systems selection by multiple attribute decision making, PhD dissertation, Kansas State University, 1981.

[41] M. Behzadian, S. K. Otaghsara, M. Yazdani, J. Ignatius, A state-of the-art survey of TOPSIS applications, Expert Systems with Applications 39 (17) (2012) 13051–13069.

[42] I. E. Grossmann, J. J. Siirola, Chemical process systems design, Dept. of Chemical Engineering, Carnegie Mellon University, http://coursecatalog.web.cmu.edu/search/?P=06-421, accessed: 01-09-2018 (2016).

[43] D. Bhattacharyya, Chemical process design, Dept. of Chemical and Biomedical Engineering, West Virginia University, https://sites.google.com/a/mix.wvu.edu/pse-deb/teaching, accessed: 01-09-2018 (2016).