Semantic–Statistical Model for Assessing Expert Competence in Scientific Competitions

Vitaliy Tsyganok[1,*,†] and Yaroslav Khrolenko[1,†]

[1] Institute for Information Recording of NAS of Ukraine, 2 Shpaka Street, Kyiv, 03113, Ukraine

Abstract

This paper addresses the problem of objectively evaluating expert competence during the formation of juries for scientific competitions. It is shown that traditional bibliometric indicators (such as the Hirsch index, number of publications, and citation counts) do not always reflect a researcher's topical relevance to the specific subject domain of a competition, particularly in cases of interdisciplinary studies. To overcome this limitation, a semantic–statistical model is proposed, based on describing the competition's thematic scope through a set of concepts and its iterative expansion via co-occurrence analysis in bibliographic databases. The expert's scientific profile is compared with a multilayer system of competition concepts, and the integral competence score is defined as a weighted sum of matches across different levels of the model.

The experimental validation of the model was carried out using data from the OpenAlex aggregator, which contains over 65,000 concepts and corpora of publications across diverse scientific domains. The study demonstrated that the model is robust to reductions in sample size: even when using only 3% of the full publication corpus, the results remain close to those of the complete model (Jaccard coefficient > 0.9). This indicates the possibility of reducing the volume of processed data without significant quality loss, thereby lowering computational requirements and enabling partial offloading of calculations to the aggregator side.

The practical significance of the proposed approach lies in the development of automated decision-support systems for organizers of scientific competitions, conferences, and grant programs. The model enhances the objectivity and transparency of expert selection, accommodates the interdisciplinary nature of research topics, and ensures interpretability of evaluation results.

Keywords

expert competence, semantic–statistical model, competitions of scientific papers, OpenAlex, conceptual units, thematic relevance

1. Introduction

In today's scientific landscape, the challenge of establishing objective and well-balanced expert committees for student research competitions, grant allocation, and project peer review is becoming increasingly significant. The quality of decisions is directly contingent upon the competence of the selected experts, as they determine whether the submitted works align with the current state of scientific advancement, demonstrate novelty, and possess practical relevance. Conventional approaches to expert selection primarily rely on bibliometric indicators—such as the Hirsch index, publication counts, citation metrics, and academic titles. Although these indicators reflect a researcher's overall scholarly contribution, they do not adequately capture contextual alignment with the specific subject area of the competition. This limitation is particularly critical in interdisciplinary research.

Emerging domains often integrate knowledge from multiple disciplines, giving rise to novel subject areas that remain insufficiently represented in traditional classification systems. Fields such

Information Technology and Implementation (IT&I-2025), November 20-21, 2025, Kyiv, Ukraine

* Corresponding author.
† These authors contributed equally.
vitaliytsyganok@gmail.com (V. Tsyganok); yaroskhr@gmail.com (Ya. Khrolenko)
0000-0002-0821-4877 (V. Tsyganok); 0009-0004-0641-827X (Ya. Khrolenko);

as bioinformatics, quantum computing, and cognitive science exemplify this trend, as they evolve at disciplinary intersections and cannot be adequately categorized through journal hierarchies or domain-specific taxonomies alone. Consequently, the selection of experts becomes problematic: strong bibliometric performance does not guarantee that a researcher's expertise is directly relevant to the competition's thematic scope. As a result, committees may include scholars with high overall qualifications but insufficient specialization to assess submissions in narrow or emerging domains.

These challenges highlight the need for models that combine semantic analysis of scientific texts with statistical methods for evaluating the co-occurrence of key concepts in global research discourse. Such an integrated approach enables a more precise assessment of thematic proximity between the competition's subject area and the expert's research profile.

2. Analysis of recent research

Over the past decades, the scientific community has proposed a variety of approaches for evaluating expert competence, each reflecting a specific aspect of a researcher's scientific activity. The most common remain formal bibliometric methods, which are based on quantitative indicators of research productivity. These include the Hirsch index, citation counts, total number of publications, as well as academic degrees and titles. Such metrics have clear advantages: they are easy to compute, well standardized, and allow for quick comparisons between researchers. However, their main drawback lies in the lack of connection to a specific subject domain. As Bornmann and Daniel [1] and Waltman and van Eck [2] have shown, formal indicators do not always correlate with actual expertise in narrow thematic fields.

To account for the content-specific nature of research, content-oriented methods have been developed. These are based on the analysis of keywords, abstracts, and descriptors accompanying scientific publications. In this case, the comparison of expert profiles and competition materials is performed by matching or measuring the frequency of recurring terms. The advantage of this approach lies in its intuitive interpretability and applicability even with small text corpora. Its weakness, however, stems from the ambiguity of natural language: the same concept can be expressed by different words or phrases, and author-provided keywords are often subjective and do not necessarily reflect the true content of the article, as emphasized by Haustein and Larivière [3].

The next stage of evolution involves semantic methods that employ natural language processing techniques to determine contextual proximity between terms. Modern vector-based models, such as Word2Vec (Mikolov et al. [4]), GloVe, or BERT (Devlin et al. [5]), represent words and concepts in a multidimensional space, where the distance between them corresponds to semantic similarity. This enables comparison not only of identical terms but also of semantically related concepts. For example, if a competition topic includes the term "machine learning," then an expert whose publications predominantly use the term "neural networks" would still be considered relevant, even without a direct keyword match.

Statistical methods, in turn, rely on studying the patterns of co-occurrence of key concepts in publications. A typical example is the construction of co-occurrence graphs, where vertices represent concepts and edges reflect their joint frequency of use. Such graphs make it possible not only to identify directly related concepts but also to study the structure of scientific knowledge at a higher level, for instance, to detect interdisciplinary links or to form thematic clusters, as discussed by van Eck and Waltman [6].

The problem of assessing expert competence correlates with the task of assigning reviewers to scientific articles. For example, Stelmakh, Shah, and Singh [7] proposed the *PeerReview4All* model, which addresses the reviewer assignment problem by maximizing topical relevance between submitted works and expert profiles. A systematic review by Zhao et al. [8] summarizes recent methods of automatic reviewer assignment, highlighting relevance and scalability as key criteria. Jovanovic et al. [9], in their review of the "Reviewer Assignment Problem," emphasize that assessing thematic competence remains central to the further development of reviewer assignment models.

Leyton-Brown, Shah, Stelmakh, and Thakur [10] describe the practical LCM system, implemented at AAAI-35, which enables scalable reviewer–paper matching based on topical scores. Similarly, Anjum et al. [11] propose a joint topical space model for aligning articles with reviewers, demonstrating high accuracy on datasets from computer architecture conferences.

In recent years, aggregator services have become widely adopted, providing large-scale access to scientific works and their associated metrics, while performing classification tasks that require the identification of conceptual units (concepts, topics, keywords), often on the basis of semantic similarity. Services such as Scopus, Web of Science, PubMed, or Google Scholar provide access to massive corpora of publications, along with bibliometric indicators (citations, h-index) and altmetric measures (social media impact, mentions). Each platform relies on its own system of conceptual classification: Scopus employs the ASJC journal classification, Web of Science uses its hierarchical subject categories, PubMed applies the MeSH vocabulary, while OpenAlex introduces a multilayered taxonomy of Fields of Study. Each of these approaches has advantages and limitations. For instance, Moed [12] notes that Scopus classification does not always reflect the actual topic of an individual article since it is tied to journal profiles. PubMed is highly effective in the biomedical domain but not applicable to other fields. OpenAlex, described by Priem, Piwowar, Orr, and Carbery [13], provides the broadest coverage, but its concepts are sometimes overly general.

Nevertheless, the numerical connectivity indicators calculated by aggregators provide information about the proximity of concepts within the entire system of scientific knowledge. However, the task of competence evaluation requires assessing the proximity of an expert's research profile to that of a specific competition. This is particularly important when competitions address emerging research areas, which often lie at the intersections of several disciplines and lack a well-established position in the broader system of knowledge.

Thus, a review of the literature shows that existing approaches can partially address the problem of expert competence evaluation, but remain either overly formal or narrowly specialized. None of them ensures accuracy, universality, and adaptability to interdisciplinary challenges simultaneously. This creates a foundation for developing hybrid models that integrate semantic and statistical features, thereby enabling a more comprehensive and relevant representation of scientific competence.

3. Problem statement

The object of this study is the process of evaluating expert competence in the context of forming a jury for scientific competitions. This process directly determines the quality of the evaluation of competition materials, since the correctness of reviewer selection affects both the objectivity and the scientific significance of the resulting decisions.

Recent research emphasizes that the key objective of reviewer assignment systems is to ensure a high degree of topical relevance between an expert's research profile and the submitted works.

The scientific problem lies in the lack of a universal competence evaluation model that would accurately capture the topical alignment of experts with the subject matter of competitions. Such a model must account simultaneously for the high degree of specialization of research topics on the one hand, and their interdisciplinary character on the other.

The aim of this work is to develop a model for evaluating the competence of jury members in scientific competitions, based on the semantic and statistical relationships between the concepts (content units) that describe the competition's subject area and the scientific works of potential experts. The proposed approach is intended to create a universal method that incorporates the interdisciplinary nature of modern scientific domains and enables the evaluation of expert competence with respect to the specific subject domain of the competition. The model should ensure completeness, accuracy, scalability, and practical applicability, while allowing adaptation to different types of content units (keywords, topics, descriptors) with minimal modifications.

4. Metodology

4.1. Description of the Competition's Thematic Through a Set of Concepts

In the context of this study, the term "concept" refers to a standardized thematic notion with a fixed identifier assigned by a scientific aggregator (such as OpenAlex, Scopus, or PubMed) as a result of the automatic classification of a publication.

Concepts possess a number of properties. Each concept is a unique element of the aggregator's controlled vocabulary (taxonomy). For example, in OpenAlex, it is an element of the Fields of Study (FoS) hierarchy, such as C41008148 ("Structural Bioinformatics"). A concept is not a free-form keyword provided by the author but the result of machine analysis of publication metadata and/or full text performed by the aggregator using NLP algorithms. The use of standardized concepts eliminates problems of synonymy (e.g., "Neural Networks" and "ANN") and polysemy (e.g., "cell" in biology vs. electrical engineering), since each semantic notion corresponds to a unique identifier. Another advantage provided by the use of aggregator platform classifiers is their multilingual capability. For instance, OpenAlex employs NLP algorithms capable of handling various languages. When processing non-English publications, the system attempts to map local terminology onto the global FoS taxonomy, ensuring comparability of works written in different languages within a single semantic space. However, the quality of classification depends on the effectiveness of processing the specific language.

A key stage in constructing a semantic–statistical model for evaluating expert competence is the formalization of the competition's subject domain. To achieve this, it is proposed to describe the competition as a set of concepts—content units that reflect its thematic scope. Concepts may be represented in the form of keywords, descriptors, topics, or other semantic markers that characterize scientific texts. Thus, the set of competition concepts forms a conceptual core, relative to which the subsequent determination of the relevance of experts' publications is performed.

The source for constructing such a set can be standardized classification systems or taxonomies used in scientific information aggregators. It is recommended to adopt the classifier of the aggregator from which the scientific works will subsequently be retrieved for analysis. The use of such classifiers allows for the standardization of the competition's thematic description and ensures its integration with bibliographic databases.

The formalization of a competition or expert review topic begins with selecting a classification system from a scientific aggregator (for example, Fields of Study in OpenAlex, ASJC in Scopus, or MeSH in PubMed), which provides access to a standardized vocabulary of concepts with unique identifiers. The initial set of competition concepts is determined through programmatic retrieval via the aggregator's public API: organizers compile a list of key thematic terms and perform search queries through the respective API. The system returns a ranked list of standardized concepts with relevance scores, which, after expert validation, form the final set of competition concepts.

In cases where the organizers of a competition cannot employ an existing classifier, or when the available system proves insufficiently detailed, the set of concepts may be formed directly through the analysis of competition documentation, abstracts, keywords, and titles of scientific works in the relevant field. For this purpose, methods of automatic keyword extraction, topic modeling (e.g., Latent Dirichlet Allocation [14]), or pre-trained language models (e.g., BERT [15], SciBERT [16]) can be applied to generate coherent sets of concepts. The construction of a classifier, however, represents a separate task that lies beyond the scope of this study.

To construct the semantic space of a competition, viewed as a subspace of the general space of scientific work concepts, a statistical–semantic approach is proposed. Within this approach, the statistical–semantic proximity of concepts is defined on the basis of the frequency of their co-occurrence in the global corpus of scientific works. Using this method, the competition model is represented as a projection of the competition's concepts onto the general model of scientific knowledge concepts (i.e., the complete body of scientific publications).

The generalized algorithm implementing this approach consists of the following steps:

Step 1. As a starting point, the set of concepts describing the competition is taken. They form the set of concepts of the initial zero level (layer):

$$T0 = \{\ t_n^0\ \}, n = \overline{1, N} \tag{1}$$

where

N – the number of competition concepts;

t_n^0– the n-th concept of the set T0.

Step 2. A sample of works W^{T0} is then selected, each of which contains at least one competition concept:

$$W^{T0} = \{\ w_i^{T0}\ \}, i = \overline{1, |W^{T0}|} \tag{2}$$

Each work in the obtained sample $\{\ w_i^{T0}\ \}, i = \overline{1, |W^{T0}|}$ is characterized by its own set of concepts $T^{w_i^{TO}}$:

$$w_i^{T0} = \langle\ T^{w_i^{TO}}\ \rangle \tag{3}$$

$$T^{w_i^{TO}} = \left\{ t_l^{w_i^{TO}} \right\}, l = \overline{1, |T^{w_i^{TO}}|} \tag{4}$$

with the restriction that the reappearance of competition concepts is not allowed:

$$T^{w_i^{TO}} \bigcap T0 \neq \emptyset \tag{5}$$

For each competition concept, the set of works is determined in which this concept occurs among their conceptual representations:

$$W^{T0} = \{\ w_i^{T0}\ \}, i = \overline{1, |W^{T0}|} \tag{6}$$

and the number q_n^0 of works in which it is used is calculated:

$$q_n^0 = |W^{t_n^0}| \tag{7}$$

Step 3. Based on the obtained set of works, a set of unique concepts is formed that are co-occurring with the concepts of the previous (zero) level. By “uniqueness” of concepts we mean not only the absence of duplicates among the identified concepts, but also the exclusion of concepts from the previous level T0. In this way, the set of first-level (layer) concepts is constructed.

$$T1 = \{t_r^1\} = \left(\bigcup_i^{|W^{T0}|} T^{w_i^{TO}} \right) \Big/ \ T0, \tag{8}$$

where $r = \overline{1, |\mathrm{T1}|}$

Step 4. On the obtained set of works W^{T0}, the number of works q_r^1 in which each concept $t_r^1 \in T1$ occurs is calculated.

$$q_r^1 = |W^{t_r^1}| \tag{10}$$

$$W^{t_r^1} = \left\{ w_i \in W^{t_r^1} \mid t_r^1 \in T^{w_i} \right\} \tag{11}$$

Based on the frequency value of concept usage, it is possible to identify and exclude random occurrences of irrelevant concepts. The threshold frequency level for excluding a concept from consideration is determined by the specifics of the subject domain and is defined by the organizers of the competition—for example, 3% of the size of the selected set of works.

$$L^1 = scope * |W^{T0}|, \qquad scope = 0.03 \tag{12}$$

The use of a 3% threshold for concept filtering is consistent with established practices in bibliometric mapping, where similar thresholds are applied to identify statistically robust and semantically meaningful relationships between concepts [17]. The purpose of this threshold is to eliminate random co-occurrences ("noise") and focus on associations that form the core of the shared scientific discourse surrounding the topic. Empirical studies have shown that thresholds in the range of 1–5% are effective for constructing balanced and interpretable models. The selected 3% value lies in the middle of this range and, as demonstrated by our experiment, provides an optimal balance between precision and completeness of the thematic model. However, the task of determining the appropriate threshold level requires further investigation.

Step 5. Similarly, on the basis of the set T1, the next sample of works W^{T1} is formed, in which at least one concept from the set T1 was used, and a new set of unique concepts T2 is obtained.

$$W^{T1} = \{ w_i^{T1} \}, i = \overline{1, |W^{T1}|} \tag{13}$$

$$T2 = \{t_r^2\} = \left(\bigcup_i^{|W^{T1}|} T^{w_i^{T1}} \right) \Big/ \left(T0 \bigcup T1 \right), \text{де } r = \overline{1, |\mathrm{T2}|} \tag{14}$$

For each concept of the set T2, the frequency of its usage q_r^2 is calculated in the sample of works W^{T1}.

$$q_r^2 = |W^{t_r^2}| \tag{15}$$

$$W^{t_r^2} = \left\{ w_i \in W^{t_r^2} \mid t_r^2 \in T^{w_i} \right\} \tag{16}$$

Concepts t_r^2, for which the frequency of occurrence q_r^2 does not satisfy the threshold level

$$L^2 = scope * |W^{T1}|, scope = 0.03 \tag{17}$$

are excluded from consideration.

Step 6. In the same way, the set of concepts of the next (n+1)-th layer is formed. The process continues iteratively until the set of concepts of the next level becomes empty.

Thus, the model is sequentially expanded (Fig.1):

$$T1 \rightarrow T2 \rightarrow . . . \rightarrow TM, \tag{18}$$

where M is the maximum depth of the model, determined either by the criterion of exhausting new concepts or by the established threshold of co-occurrence frequency. As a result, the application of this algorithm yields a model of the competition in which concepts are grouped according to levels of semantic connectivity. The connectivity of concepts is constructed on the basis of statistics of their co-occurrence in scientific works from the global corpus of research.

The distinctive feature of the proposed model is that it constitutes a system of organizing scientific knowledge, in which concepts are arranged according to their increasing semantic distance (layer/level number) from the conceptual core defined by the competition's thematic scope. In other words, the model implements a procedure of structural localization of general scientific knowledge with respect to the coordinate system represented by the set of competition concepts.

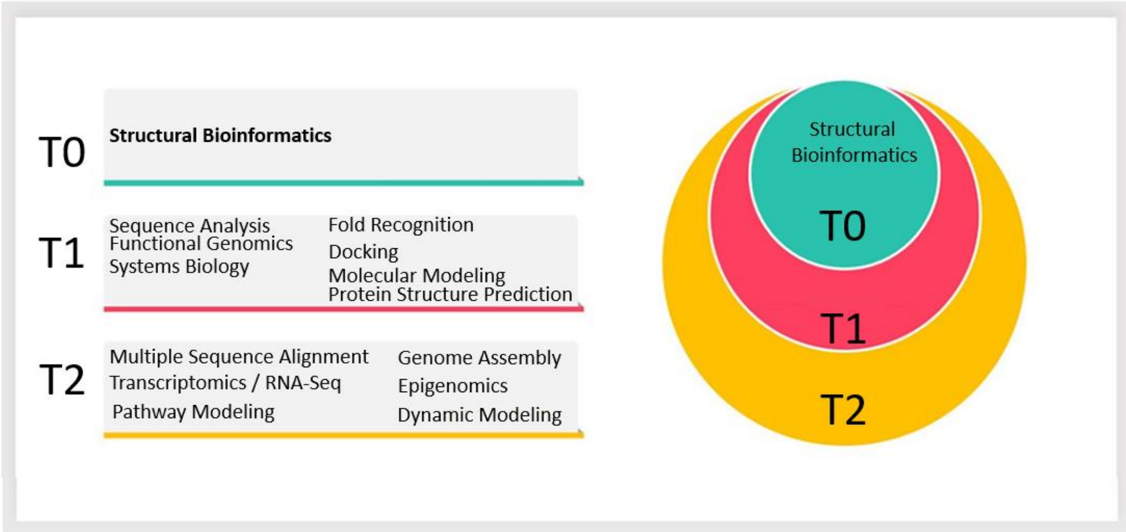

Figure 1: Example of ordering by increasing semantic distance for the concept "Structural Bioinformatics"

4.2. Properties of the Model

The proposed semantic–statistical model has a number of properties that determine its suitability for use in systems for forming juries of scientific competitions.

4.2.1. Completeness

The model ensures the gradual coverage of the entire relevant subject domain due to its iterative construction principle. At each step, sets *Tk* are formed, which expand the competition core *T0* with new concepts co-occurring in the corpus of scientific publications. This makes it possible to identify not only explicitly stated terms but also related concepts that frequently appear in the same contexts as the competition's concepts. Thus, the model is not limited to a narrow list of keywords but captures a broader spectrum of notions that genuinely reflect the structure of knowledge in the chosen field.

4.2.2. Universality

The construction methodology is independent of the specific type of content units. Concepts may take the form of keywords, descriptors, topics, MeSH terms, Fields of Study (FoS) from OpenAlex, or even phrase vectors generated using transformer-based models (e.g., BERT, SciBERT). This means that the model is easily integrated with various databases (Scopus, Web of Science, PubMed, OpenAlex) and can be applied in biomedical research as well as in engineering or the humanities. Its universality makes it a flexible tool for organizers of competitions with different profiles.

4.2.3. Scalability

The model is applicable to datasets of any size, including large-scale data. The use of statistical methods (once a relevant sample is obtained) does not require processing excessively large volumes and allows for a significant reduction in computational and storage requirements for intermediate results. Since at each iteration of the algorithm the concepts used in previous steps are excluded, the overall size of the model cannot exceed the size of the aggregator's concept dictionary. For example, in the OpenAlex database the size of such a dictionary is approximately 65,000 concepts across all scientific disciplines.

4.2.4. Objectivity

Unlike traditional evaluation methods based on bibliometric indicators (e.g., Hirsch index, number of publications), the proposed model evaluates thematic relevance directly. It reduces the impact of subjective factors, since it relies on formalized metrics such as co-occurrence frequency of concepts and distances in vector space. As a result, an expert whose research profile strongly correlates with the competition's subject area receives a high score regardless of formal status or citation counts. This enhances the fairness of jury formation from a scientific and methodological perspective.

4.2.5. Interpretability

The model enables transparent explanation of results. Each competence evaluation can be detailed by indicating which concepts of the expert's profile overlap with the competition core and which belong to more distant levels. This allows organizers to justify why a particular expert is considered relevant and to compare experts based on clear, understandable criteria.

4.2.6. Flexibility

The algorithm allows for the adjustment of weighting coefficients α_k which regulate the sensitivity of the model to more distant concepts. If a strict evaluation focused only on the competition core is required, rapidly decaying coefficients may be used ($\alpha_k = 1/_{2k}$). If broader context needs to be considered, slower decay may be applied ($\alpha_k = 1/_{(k+1)}$. This makes the model adaptable to different types of competitions and subject domains.

4.2.7. Accuracy

Theoretical accuracy (accuracy by design).
The model is constructed so that the core T0 carries the maximum weight, while distant levels (T1, T2, …) gradually lose influence due to decreasing coefficients α_k . This makes the system robust to "noisy" concepts and ensures that the main contribution to the final evaluation comes from thematically relevant notions. In other words, the model is theoretically accurate, provided that the competition core is adequately defined.

Practical accuracy (evaluation on data).

The accuracy of the model is determined by its ability to correctly identify experts whose research profiles truly correspond to the competition's subject domain. It depends on the quality of the initial set of concepts T0, the completeness of the expert publication corpus, and the classification system used for content units.

A drawback of the proposed model can be considered the fact that, in the pursuit of resource efficiency and effectiveness, it has lost its autonomy. It requires extensive and labor-intensive data preprocessing. In particular, this involves assigning thematic concepts to scientific works based on their titles, abstracts, keywords, and textual content, or constructing dictionaries of key phrases, topics, and so forth. However, this drawback cannot be regarded as critical, since many modern aggregators of scientific publications already perform such processing and provide the results in open access.

4.3. Integral Evaluation of Expert Competence

The constructed semantic–statistical model provides for the formation of a generalized numerical indicator that reflects the degree of correspondence between an expert's scientific profile and the thematic scope of the competition. Within the framework of the problem being addressed, the scientific profile of an expert is represented as the complete list of their publications.

The integral indicator of expert competence is defined as a weighted sum of matches between the concepts from the expert's scientific works and the concepts of the competition's thematic model.

4.3.1. Model

Each expert has a set of scientific works $W(E)$.

Each work of the expert $\left\{ w_j \right\}$ is described by a set of concepts $\{ p^{W_j} \}, j = \overline{1, |\mathrm{W(E)}|}$

The expert profile E is represented by the set of unique concepts derived from their works:

$$P(E) = \{p_1, p_2, \dots, p_V, \} \tag{19}$$

$$p_i \in \left(\bigcup_{j=1}^{|W(E)|} \{ p^{W_j} \} \right), \tag{20}$$

where

p_i is a concept present in the expert's works;

V is the total number of unique concepts.

To eliminate any ambiguity regarding the formation of an expert's scientific profile P(E), it should be emphasized that the set of unique concepts associated with an expert is derived exclusively from the classification provided by the selected scientific aggregator and does not involve the use of any proprietary extraction algorithms.

This approach ensures that the expert's profile is constructed within the same semantic space (e.g., the OpenAlex Fields of Study space) as the competition's thematic model T, which is a necessary condition for the correct computation of the integrated evaluation score $C(E)$. The methodology is transparent and reproducible, as it relies entirely on publicly available data and metadata supplied by the aggregator.

The set of competition concepts T is represented as a multilayer structure:

$$\mathrm{T0}, \mathrm{T1}, \dots, \mathrm{TM} \tag{21}$$

where *T0* is the competition core, and each subsequent level Tk includes concepts statistically associated with the previous level.

To reflect the varying significance of levels in the competition model, weighting coefficients $\alpha_k$were introduced, which decrease as the distance from the core increases:

$$\alpha_1 > \alpha_2 > \cdots > \alpha_L \tag{22}$$

A typical choice is geometric decay:

$$\alpha_k = \frac{1}{2^k} \tag{23}$$

4.3.2. Matching Function

For each level Tk, a matching function is defined:

$$f(E, Tk) = \sum_{c \in Tk} 1\,[c \in P(E)] \cdot q_E(c) \tag{24}$$

where

$1\,[c \in P(E)]$— an indicator of the presence of concept c in the expert's profile,

$q_E(c)$ — the weight of concept ccc in the expert's profile (the number of works in which concept ccc occurs).

Thus, $f(E, Tk)$ reflects the intensity of the presence of competition concepts from level Tk in the expert's research output.

The integral evaluation of expert competence is defined as:

$$C(E) = \sum_{k=0}^{L} \alpha_k \cdot f(E, Tk) \tag{25}$$

This indicator generalizes the matches across all levels of the model, assigning the greatest weight to the concepts of the competition core while reducing the influence of more distant terms.

4.3.3. Normalization

A direct interpretation of the obtained evaluation results for an individual expert has little meaning without comparison to the results of other experts in the group. Therefore, it is reasonable to normalize the evaluation results:

$$C_{norm}(E) = \frac{C(E)}{max_{E'}\, C(E')} \tag{26}$$

where the maximum is taken over the group of all experts E′.

This normalization allows the evaluation to be interpreted on a relative scale [0;1].

4.3.4. Interpretation of Results

- High values $C_{norm}(E) \geq 0.7$ indicate strong topical relevance of the expert's profile to the competition's subject area.
- Medium values $0.4 \leq C_{norm}(E) < 0.7$ reflect partial relevance, which may be acceptable in interdisciplinary competitions.
- Low values $C_{norm}(E) < 0.4$ signal that the expert's research activity is distant from the competition's subject area.

5. Experiment

5.1. Objective and Hypothesis

The objective of the experiment is to test the hypothesis that, for constructing a valid semantic–statistical model of the competition's thematic scope, the use of the entire corpus of scientific publications is redundant. It is sufficient to limit the analysis to a "relevant sample" of a certain size. Thus, the task of the experiment is to determine the minimal sample sizes that guarantee the stability of the model.

5.2. Input Data

For the experiment, data from the scientific information aggregator OpenAlex were used. The study employed:

- A corpus of scientific publications in the field of «Computer Science» containing 2 million works.
- Publication metadata.
- The OpenAlex concept directory (approximately 65,000 concepts) [18].

5.3. Experimental Scenario

- Initial data formation. A random set of concepts with varying degrees of specialization is selected (from different levels of the hierarchy in the OpenAlex concept taxonomy).
- Sampling of works. For each selected concept, a set of scientific works in which it appears is formed (see Sec. 4.1, Step 2 of the algorithm). The number of such works is determined (see Table 1, column "Number of Papers in Full Sample").
- Formation of co-occurring concepts. Based on the obtained set of works, the set of unique concepts co-occurring with the studied concept is computed (see Sec. 4.1, Step 3 of the algorithm).
- Frequency estimation. For each co-occurring concept, the number of works in which it appears is counted, and its percentage relative to the sample size is calculated (see Sec. 4.1, Step 4 of the algorithm).
- Noise filtering. Concepts with a frequency below 1% (scope=0.01scope = 0.01scope=0.01) are excluded from the set to eliminate random occurrences of irrelevant terms and informational noise.
- Obtaining the set of relevant concepts. As a result, a subset of concepts co-occurring with the selected concept within the publication database is formed.
- Artificial reduction of the sample. The number of works from Step 2 is intentionally reduced to 30%, 3%, and 0.1% of the full size.
  For each reduced sample, Steps 3–6 are repeated.
- Evaluation of similarity of results. The sets of concepts obtained from the full and reduced samples are compared using the Jaccard coefficient:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (27)$$

The Jaccard coefficient measures the similarity between sets and is defined as the size of their intersection divided by the size of their union.

5.4. Results

Table 1 presents the values of the Jaccard coefficient for selected concepts with varying degrees of specialization.

Table 1
Values of the Jaccard coefficient for different sample sizes across concepts with varying levels of specialization

Concept	Number of Papers in Full Sample	Jaccard Coefficient		
		30% of Full	3% of Full	0.1% of Full
Mathematics	313644	1	1	1
Software engineering	30910	1	0,933333333	0,757575758
Artificial intelligence	369553	1	1	0,968253968
Artificial neural network	130556	1	1	0,982758621

The analysis of the obtained data shows that the semantic–statistical model of the competition's thematic scope is highly robust to reductions in sample size. Even with a substantial decrease in the number of works (down to 3% of the full corpus), the resulting sets of concepts remain close to those constructed using the complete dataset. Significant deviations are observed only for very small samples (0.1%), particularly in the case of rare or highly specialized concepts.

Practical implications of the obtained results include:

- the possibility of significantly reducing the volume of processed data without substantial loss of model quality;
- reduced computational requirements for implementing the algorithm;
- potential offloading of the most resource-intensive computations to the aggregator side.

6. Discussion

The proposed method evaluates the overall proximity of the concepts in an expert's works to the competition's thematic scope as a generalized abstract entity. However, in practice, the actual competition scope may combine several different directions within one discipline, or even be interdisciplinary. In such cases, to obtain a correct evaluation, it is necessary to involve experts from multiple domains. This is confirmed by established practices in forming expert groups. Another option is to subject each work to independent evaluation by several experts, followed by weighting of the assigned scores with respect to the experts' competence indices. Within these approaches, the lack of competence of a particular expert is compensated by the competences of other members of the group.

From this perspective, the proposed method requires further development toward differentiating the evaluation for each competition concept individually. This, in turn, enables the formation of expert groups using multi-criteria optimization methods.

7. Conclusion

The developed model for evaluating the competence of jury members in scientific competitions, based on a semantic–statistical approach, provides a universal and objective method for selecting experts, with adaptability to interdisciplinary and applied domains. The use of semantic models and statistical methods of co-occurrence frequency analysis makes it possible to construct a comprehensive and scalable representation of the competition's subject domain, reflecting semantic proximity between competition concepts and the expert's scientific works.

Integration with aggregator services provides access to large volumes of up-to-date and structured data (scientific articles with performed terminological analysis and concept identification, dictionaries of concepts, key phrases, and topics), which enables the practical implementation of the method and ensures its effectiveness.

The proposed algorithm accounts for both direct and indirect connections by forming a hierarchy of semantic connectivity levels. The integral competence evaluation of an expert is based on the intersection of concepts from their scientific works with the levels of the proposed semantic–statistical model of the competition's thematic scope. The evaluation incorporates semantic distance, reducing the contribution of concepts from distant levels in a geometric progression, thus ensuring convergence of the evaluation measure.

The method is flexible and adaptable to different types of content units. In addition to concepts, it can incorporate keywords, topics, and descriptors, making it compatible with diverse approaches to structuring scientific knowledge.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT-4 and Grammarly in order to: Grammar and spelling check. After using these tool(s)/service(s), the authors reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] L. Bornmann and H.-D. Daniel, "What do citation counts measure? A review of studies on citing behavior," *Journal of Documentation*, vol. 64, no. 1, pp. 45–80, Jan. 2008. doi: 10.1108/00220410810844150.

[2] L. Waltman and N. J. van Eck, "Field-normalized citation impact indicators and the choice of an appropriate counting method," *Journal of Informetrics*, vol. 9, no. 4, pp. 872–894, Oct. 2015. doi: 10.1016/j.joi.2015.08.001.

[3] S. Haustein and V. Larivière, "The use of bibliometrics for assessing research: Possibilities, limitations and adverse effects," in *Incentives and Performance: Governance of Research Organizations*, I. M. Welpe, J. Wollersheim, S. Ringelhan, and M. Osterloh, Eds. Cham: Springer, 2015, pp. 121–139. doi: 10.1007/978-3-319-09785-5_8.

[4] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, Jan. 2013. [Online]. Available: https://arxiv.org/abs/1301.3781.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, Minneapolis, MN, USA, 2019, pp. 4171–4186. doi: 10.48550/arXiv.1810.04805.

[6] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, Aug. 2010. doi: 10.1007/s11192-009-0146-3.

[7] I. Stelmakh, N. Shah, and A. Singh, "PeerReview4All: Fair and accurate reviewer assignment in peer review," *Journal of Machine Learning Research*, vol. 22, no. 142, pp. 1–54, 2021. [Online]. Available: http://jmlr.org/papers/v22/20-1061.html.

[8] H. Zhao, Y. Zhang, Y. Fang, X. Hu, and J. Tang, "Automatic academic reviewer assignment: A survey of the state-of-the-art," *ACM Transactions on Intelligent Systems and Technology*, vol. 13, no. 5, pp. 1–27, Sep. 2022. doi: 10.1145/3531046.

[9] J. Jovanovic, B. Stantic, and C. Wagner, "Reviewer assignment problem: A survey of the state of the art," *Artificial Intelligence Review*, vol. 56, pp. 3685–3724, 2023. doi: 10.1007/s10462-022-10244-0.

[10] K. Leyton-Brown, N. Shah, I. Stelmakh, and M. Thakur, "The learning-to-match (LCM) system for reviewer assignment: Design and deployment at AAAI-35," *AI Magazine*, vol. 45, no. 1, pp. 50–61, Mar. 2024. doi: 10.1002/aaai.12139.

[11] M. Anjum, D. Choudhury, and T. B. Dinesh, "Learning a joint topical space for paper and reviewer representation towards automatic paper-reviewer assignment," in *Proc. ACM/IEEE Joint Conf. Digital Libraries (JCDL)*, Champaign, IL, USA, 2019, pp. 185–194. doi: 10.1109/JCDL.2019.00037.

[12] H. F. Moed, "Citation analysis in research evaluation," *Information Science and Knowledge Management*, vol. 9, Springer, 2005. doi: 10.1007/1-4020-3714-7.

[13] J. Priem, H. Piwowar, R. Orr, and S. Carbery, "OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts," *arXiv preprint arXiv:2205.01833*, May 2022. [Online]. Available: https://arxiv.org/abs/2205.01833.

[14] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, Jan. 2003. [Online]. Available: https://jmlr.org/papers/v3/blei03a.html.

[15] Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, Minneapolis, MN, USA, 2019, pp. 4171–4186. doi: 10.48550/arXiv.1810.04805.

[16] Beltagy, K. Lo, and A. Cohan, "SciBERT: A pretrained language model for scientific text," in *Proc. EMNLP-IJCNLP*, Hong Kong, China, 2019, pp. 3613–3618. doi: 10.48550/arXiv.1903.10676.

[17] Waltman, L., van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629–635. doi: 10.1016/j.joi.2010.07.002

[18] https://api.openalex.org/concepts