This study examines multi-context educational data acquired from independent digital platforms and characterized by differences in the composition of entities, types of interactions, and data representation schemes. The task addressed relates to the lack of a unified model for the invariant representation of heterogeneously structured data, the assessment of their structural compatibility, the determination of the costs of mutual adaptation, and the possibility of comparative analysis of digital platforms while preserving the characteristics of the educational context.

A parametrically typed invariant hypergraph model has been proposed, which represents educational data as typed hypergraphs, their semantic normalization into a single type space, as well as the calculation of indicators of structural invariance, adaptation, and integral utility. A parametric analysis of sensitivity to changes in the data structure and model weighting coefficients has been suggested.

The novelty of this work is underscored by the following:

– a unified parametric hypergraph model has been constructed that combines the invariant representation of educational data, a quantitative assessment of the structural compatibility, adaptation costs, and integral utility of digital educational platforms;

– a parametric analysis has been proposed to examine the influence of structure and weighting factors on the integration and comparative analysis of platforms;

– the model was experimentally tested on data from independent educational platforms, including over 3,100 educational records and 4,861 hyperedges, revealing differences in the structural compatibility of the studied contexts.

The model supports decision making in the selection and design of digital educational ecosystems in the context of structurally heterogeneous data

Keywords: invariant hypergraph, data integration, multi-context educational data, hypergraph learning

UDC 004.048

DOI: 10.15587/1729-4061.2026.369576

CONSTRUCTION OF A PARAMETRIC INVARIANT HYPERGRAPH MODEL FOR INTEGRATING AND ANALYZING MULTI-CONTEXT EDUCATIONAL DATA

Kainizhamal Iklassova
PhD, Associate Professor
Department of Information and Communication Technologies*
ORCID: https://orcid.org/0000-0002-8330-4282
Aigul Shaikhanova
PhD, Associate Professor**
ORCID: https://orcid.org/0000-0001-6006-4813
Gulnar Kim
Corresponding author
PhD, Associate Professor
The International Campus*
E-mail: gakim@ku.edu.kz
ORCID: https://orcid.org/0000-0001-8402-8323
Kanat Myrzakhanov
Head of Marketing
CareerHub.kz LLP
Samal-3 Neighborhood, 15, Building 3, Almaty, Republic of Kazakhstan, 050000
Kazakh-British Technical University
Tole bi str., 59, Almaty, Republic of Kazakhstan, 050000
ORCID: https://orcid.org/0009-0000-0584-4067
Abduraimova Bayan
Candidate of Science**
ORCID: https://orcid.org/0000-0003-3913-1895
Madina Bazarova
PhD, Associate Professor
Department of Computer Modeling and Information Technology
Sarsen Amanzholov East Kazakhstan University
30th Guards Division str., 34, Ust-Kamenogorsk,
Republic of Kazakhstan, 070002
ORCID: https://orcid.org/0000-0003-2580-6580
*M. Kozybayev North Kazakhstan University
Zhumabayeva str., 114, Petropavlovsk, Republic of Kazakhstan, 150000
**Department of Information Security
L.N. Gumilyov Eurasian National University
Satbayev str., 2, Astana, Republic of Kazakhstan, 010000

Received 18.05.2026
Received in revised form 31.07.2026
Accepted date 10.08.2026
Published date 28.08.2026

How to Cite: *Iklassova, K., Shaikhanova, A., Kim, G., K. Myrzakhanov, A. Bayan, M. Bazarova (2026). Construction of a parametric invariant hypergraph model for integrating and analyzing multi-context educational data. Eastern-European Journal of Enterprise Technologies, 4 (2 (142)), 59–69. https://doi.org/10.15587/1729-4061.2026.369576*

1. Introduction

In modern information technology, special attention is paid to the digitalization and automation of data collection, processing, and analysis. In this context, the issues of integrating and comparing heterogeneous data coming from independent sources are particularly significant. In education, such tasks associated with combining data from multiple

Copyright © 2026 Authors. This is an open access article under the Creative Commons CC BY license

independent digital platforms are a key example in the field of hypergraph learning, heterogeneous data integration, and educational analytics.

The simplest example of such a problem can be stated as follows:

– there is a set of independent digital educational platforms, each of which records learner interactions in its own internal data structure;

– the data structure of each platform is specific: it uses its own entity schemas, resource identifiers, and interaction categories;

– educational interaction can simultaneously unite more than two participants or objects (learners, resources, competencies, contexts), i.e., it is polyadic rather than paired;

– data from all platforms must be represented in a single formal space without losing the structural information specific to each context;

– it is necessary to quantify the degree of structural similarity between platforms and the complexity of their mutual integration;

– the comparison results must yield a single comparative indicator of usefulness applicable to management decision-making.

This problem statement has numerous practical applications. For example, hypergraph representation is already being used to integrate heterogeneous data sources in other subject areas – for example, a hypergraph model for integrating multi-omics data enables the combination of several types of biomedical measurements into a single representation for the classification and identification of biomarkers [1]. A similar issue arises when selecting digital educational platforms for implementation at educational institutions, when designing national and inter-university educational ecosystems, and when assessing the readiness of heterogeneous information systems for integration.

A basic solution to the problem of representing heterogeneous data has been proposed within the framework of hypergraph learning theory [2]. It has been shown that a hypergraph, unlike a classical graph, which describes only pairwise relationships, allows for grouping an arbitrary number of objects into a single relationship, making it a natural tool for representing the multilateral interactions characteristic of the educational process. The active evolution of this field is confirmed by the growing number of review studies on methods for representing hypergraphs and heterogeneous graphs [3–5], as well as on methods for modeling and integrating knowledge graphs [6].

With the evolution of multi-platform digital educational ecosystems, the need arises for new approaches to the quantitative comparison of disparate data sources. The institutionalization of this area is evidenced, in particular, by the formation of a separate methodological framework for learning analytics as an independent field of research [7]. Official reviews note that modern educational organizations simultaneously use a significant number of independent digital tools, while the lack of data exchange mechanisms between them leads to the re-entry of information and the accumulation of unused data at the system level [8]. This makes it urgent not only to formally combine data but also to quantitatively assess the structural compatibility of platforms, the cost of their mutual adaptation, as well as their comparative usefulness in terms of the educational process.

Given the increasing complexity of digital educational ecosystems, the task of integrating data from independent platforms with disparate representation structures is particularly pressing. These differences complicate the shared use of educational data and its comparison across digital contexts. Therefore, devising methods for the unified representation and quantitative analysis of the structural compatibility of multi-context educational data is a pressing scientific challenge. The solution to this problem is necessary to improve the efficiency of data integration and analytical support for the management processes within digital educational ecosystems.

2. Literature review and problem statement

Hypergraphs as a tool for analyzing heterogeneous data are used in various applied tasks, including visual analytics of complexly structured datasets, where hypergraph representations enable the identification of hidden group dependencies between heterogeneous object attributes [9]. This idea is further developed in [2], in which a hypergraph neural network (Hypergraph Neural Network) is proposed that generalizes convolutional graph models to structures of arbitrary connectivity. It is shown that representing data as a hypergraph enables the identification of stable group dependences between objects that are inaccessible using traditional pairwise graph models. However, the proposed model focuses on the analysis of data initially represented as a single hypergraph and does not consider the task of comparing multiple independently constructed hypergraph structures.

Similarly, a generalized spectral-theoretic framework for hypergraph learning is proposed in [10] for clustering, classification, and object embedding problems. It has been shown that considering group interactions, rather than just pairwise interactions, improves the performance of classification tasks compared to classical graph-based methods. However, issues related to assessing the structural similarity between hypergraphs constructed from data of different natures and origins remain unresolved. This may be because the models under consideration were initially built for analyzing a single data source, and formalizing the measure of representation invariance across disparate sources has not been the subject of research to date.

One way to overcome this difficulty is to use the hypergraph directly as a mechanism for integrating multiple heterogeneous data sources, rather than simply as a means of representing a single source. This approach was used in [1], in which a hypergraph network for integrating multi-omics biomedical data (mRNA, DNA methylation, microRNA) was proposed. It was shown that the hypergraph can simultaneously reflect correlations both within and between heterogeneous data modalities, which improves classification accuracy and enables the identification of biomarkers inaccessible through separate modality analysis. However, this approach is designed for a strictly fixed set of omics modalities with a predetermined biological nature and does not address the problem of integrating data from independent, independently administered information systems, such as digital educational platforms.

Hypergraph models have also been constructed in the field of education. In [11], the EduCross framework was proposed, combining adversarial learning with a bipartite hypergraph representation for cross-modal search in multimodal educational presentations. It has been shown that the use of a hypergraph representation enables efficient com-

parison of textual and graphical components of educational content. However, this model addresses the task of matching modalities within a single collection of materials and does not address the integration of student behavioral data across different educational platforms.

Another approach to overcoming structural data heterogeneity is reported in [12], in which the Heterogeneous Graph Transformer (HGT) model is proposed. It uses a type-dependent attention mechanism to process graphs with heterogeneous vertex and edge types within a single structure. This approach is shown to provide a significant improvement in the quality of problem solving compared to existing graph models and can be scaled to web-scale graphs. However, this approach was tested within a single graph (an academic network), while the task of matching several fundamentally independent information systems with their own data schemas was not considered in this work. This limitation may be due to the fact that the combination of vertex and edge types in HGT is performed within a pre-determined, unified graph schema, whereas such a common schema is initially absent for independent educational platforms.

The task of matching schemas of heterogeneous data sources in the absence of a common schema is considered in the field of ontological alignment. In [13], methods for automatically finding matches between heterogeneous ontologies and data schemas are systematized. It is shown that the use of structural, lexical, and semantic matching methods allows for partial automation of the process of integrating heterogeneous schemas. This approach is further developed in [14], in which a review of knowledge graph refinement methods is given, demonstrating that the quality of automatically constructed matches between schemas depends significantly on the completeness and consistency of the source data. However, the authors of both works note that fully automatic and reliable matching of large-scale heterogeneous schemas remains a labor-intensive task, making the practical application of such methods challenging for dynamically changing multi-platform educational data.

The need for a comprehensive analysis of digital learning data is substantiated in [15], which demonstrates that the growing volume of data generated by digital educational environments creates a need to devise learning analytics methods to identify patterns in the educational process. The cited work substantiated the relevance of this area as a whole but did not address the issues of formal comparison of structurally heterogeneous sources of educational data. The evolution of this area is provided in [16], which gives a review of methods for the intelligent analysis of educational data and learning analytics. It is shown that the sources, formats, and analytical tasks associated with educational data are more diverse than ever before; however, most of the methods considered are focused on data analysis within a single educational system. The tasks of quantitatively comparing the structural compatibility of several independent platforms remain poorly understood.

Thus, our review of the literature [1, 2, 9–16] shows that existing hypergraph models and schema matching methods do not provide a unified formal framework. Such a framework must simultaneously provide an invariant representation of multi-context educational data, a quantitative assessment of their structural compatibility, and an estimate of the cost of integration between independent digital educational platforms. All of this makes it advisable to conduct research on the construction of a parametric invariant hypergraph model for the integration and analysis of educational data from various contexts.

3. The aim and objectives of the study

The goal of our work is to build a parametric invariant hypergraph model that provides a unified representation, quantitative comparison, and compatibility analysis of multi-context educational data. This will enable educational organizations and platform developers to compare disparate educational systems, reduce integration costs, and ensure informed selection of educational platforms based on quantitative metrics of structural compatibility and utility.

To achieve this goal, the following tasks are set:

– to formalize a typed hypergraph to represent the invariant structure of educational data representation;

– to propose quantitative metrics of structural invariance, adaptation cost, and the integrated utility of educational contexts;

– to develop a parametric modeling framework that enables analysis of the impact of structural perturbations and weighting coefficients;

– to conduct a computational experiment on real educational data from three independent educational platforms;

– to evaluate the effectiveness of the model by analyzing the structural compatibility, integration cost, and comparative utility of the educational platforms.

4. The study materials and methods

The object of our study is multi-context educational data acquired from independent digital platforms and characterized by differences in the composition of entities, types of interactions, and data representation schemes. Multi-context educational data refers to heterogeneous educational data acquired from several independent digital educational contexts. These contexts include platforms, courses, and institutional environments, and must be brought into a single analytical space for subsequent structural comparison.

The principal hypothesis assumes that the possibility of bringing structurally heterogeneous educational data into a single representation does not imply their structural identity. It is assumed that the degree of structural compatibility of educational contexts can be quantitatively assessed based on the characteristics of the composition of educational entities, interactions, and the costs required to transition between contexts.

The following assumptions were adopted in conducting the study:

– the semantic categories of educational entities and interactions are defined before the start of the computational experiment and remain unchanged within each experimental scenario;

– initial platform-specific identifiers are not considered a prerequisite for structural correspondence between data across different platforms;

– structural matching of educational contexts can be performed based on the semantic types of entities and interactions without the mandatory matching of initial platform identifiers.

The following simplifications were accepted for conducting the computational experiment:

– semantic matching of entities between platforms is performed based on predefined normalization rules; automatic construction of such correspondences is not performed within the framework of this study;

– the weighting coefficients used in calculating the model's composite indicators are determined expertly and are not automatically optimized in the baseline experiment;

– the impact of changes in model parameters is studied through individual parametric sensitivity analysis scenarios.

Before conducting the computational experiment, the source data was de-identified and verified for completeness of platform, user, and educational institution identifiers.

The computational experiment was conducted using a dataset containing 3,141 educational records obtained from three digital educational platforms: Assessment, Twig, and BilimLand (Table 1). BilimLand and Twig are part of the Bilim Group educational ecosystem, with Twig being the licensed Kazakhstani version of the Twig Education service (UK). The Assessment platform is used to conduct diagnostic assessments of students.

The data was provided by Bilim Group upon official request for the purposes of this study and is not a publicly available (open) dataset. The download contains de-identified information: students and educational institutions are represented by numeric identifiers without last names, first names, or other personal information. Records with incomplete or inconsistent platform, user, or educational institution identifiers were excluded from the analysis.

Table 1

Input data for educational platforms

Context	Records
Assessment	1000
Twig	1507
BilimLand	634
Total	3141

The computational implementation of the model is performed in the Python 3.x programming language in the Google Colab (Google LLC) cloud computing environment, which does not impose special requirements on local equipment: the calculations are reproducible on a standard CPU-runtime configuration without the use of a GPU. To process data and build hypergraph structures, the libraries pandas and numpy (loading, cleaning, and aggregation of tabular data), networkx (representation and visualization of graph structures), matplotlib, and seaborn (visualization of results), scipy (statistical calculations and parametric sensitivity analysis), openpyxl (reading source files in .xlsx format) were used.

The source code is implemented in the form of a reproducible Colab notebook with a separate parameterized block of settings (weighting coefficients α, β, γ, λ_1, λ_2, λ_3, thresholds min_entity_count and min_hyperedge_size, perturbation parameter p, initial value of the random number generator random_seed), which allows one to re-run the experiment with other parameter values without changing the main code (Table 7).

The flexibility component F remained constant across all the contexts examined and contributed little to platform differentiation. The educational performance indicator combines heterogeneous metrics from different platforms, requiring more complex semantic normalization procedures. This study examined only three educational platforms within a single ecosystem; validating the model on additional LMS (Learning Management Systems), MOOCs (Massive Open Online Courses), and mentoring systems would improve the generalizability of the results. Automatic semantic matching of entities between platforms remains beyond the scope of this study; the development of methods for matching ontologies and knowledge graphs offers prospects for its further automation.

5. Results of the construction of a parametric invariant hypergraph model for the integration and analysis of multi-context educational data

5. 1. Results of the formalization of a typified educational hypergraph

A multi-context educational environment is represented as a family of local contexts

$$C=\{C_1,C_2,\ldots,C_m\}, \tag{1}$$

where each context corresponds to a specific educational platform, a specific instance of a course, an institutional subsystem, or an analytical section of the learning process.

For each context C_i, a local hypergraph structure [17] is assigned, in the form

$$H_k=(V_k,E_k,\tau_k,\sigma_k),\ \ k=1,\ldots,m, \tag{2}$$

where V_i is the set of vertices, E_k is the set of hyperedges

$$\tau_k:V_k\to T_V, \tag{3}$$

is a function of the vertex typing, and the expression

$$\sigma_k:E_k\to T_E, \tag{4}$$

is a hyperedge typing function. The set of node types T_V includes invariant semantic categories rather than platform-dependent data fields [18]. For example, in an educational context, such categories might be student *S*, teacher *I*, learning object *LO*, competency or concept *C*, assessment event *A*, activity trace *AT*, timestamp *TM*, and organizational unit *OU*.

Let the set of hyperedge types T_E define a set of permissible patterns of educational interactions, such as:

– participation in assessment, access to an educational resource;

– assignment completion;

– competency development recording;

– teacher-mediated interaction. By separating semantic types from specific data storage schemes, the model follows the logic of knowledge hypergraphs. They preserve the high-order structure of relations better than binary graphs [19], transferring this principle to the problem of integrating educational data rather than to abstract retraining of knowledge bases. The hyperedge will represent the local educational interaction through the following expression

$$e=\{v_1,v_2,\ldots,v_r\},\ \ r\geq 2, \tag{5}$$

which allows for the unification of all entities involved in a single educational act. For example, a situation in which a learner completes an assessment task on the platform at a specific point in time, thereby demonstrating mastery of a

certain competency, can be represented by a hyperedge with the signature {*S*, *A*, *C*, *TM*}. Interaction with a learning resource can be specified by the signature {*S*, *LO*, *C*, *TM*}. This thus allows for the capture of polyadic relationships that are systematically lost when reducing data to pairwise graph relationships. When context changes, this preserves valid signatures of higher-order interactions for the analysis of both systems.

Global invariant hypergraphs are defined as an abstract structure formalizing the following statement

$$H^* = \left(T_V, T_E, \Sigma, \Pi\right). \quad (6)$$

Here, T_V and T_E are global sets of semantic types of nodes and hyperedges, Σ is the set of admissible type signatures of hyperedges, and Π is the set of projection rules. This representation allows mapping local raw data schemas into invariant semantic types [20]. Each considered local context C_k induces a hypergraph H_k through a context-dependent instantiation function represented in accordance with the following expression

$$\Phi_k : D_k \to H_k, \quad (7)$$

Here, D_k denotes the raw, unprocessed transactional data of the corresponding context. The role of the variable Π in this regard is crucial, as it is this set of rules that defines the semantic normalization layer through which heterogeneous logs from the Learning Management System or other platforms are translated into a common vocabulary of entities and interactions.

An invariant representation of educational data is understood as a unified, typified hypergraph form for describing educational entities and interactions, formed after semantic normalization, regardless of the characteristics of the original digital platform. This representation is a property of the proposed model and ensures the comparability of multi-context educational data in a single analytical space. In contrast, structural invariance characterizes the degree of similarity between two educational contexts and is considered a quantitative characteristic determined based on the SI index. Thus, an invariant representation is the result of data transformation, while structural invariance is a measurable property of the educational contexts being compared.

The formalization of a typed educational graph was performed using the initial data, the characteristics of which are given in Table 1. The largest share of records comes from the Twig platform, containing approximately 48% of all observations, while BilimLand accounts for the smallest share of the dataset. This imbalance reflects differences in student activity rather than inconsistencies in data collection and, therefore, represents a realistic multi-platform educational environment.

The construction of the typed educational hypergraph resulted in the identification of 2,509 unique nodes, distributed across five semantic categories, as shown in Table 2.

Table 2

Distribution of hypergraph vertices by categories

Object	Quantity
Students	2374
Educational institutions.	77
Competencies	37
Educational objects	20
Grades	1
Total	2509

Students account for 94.6% of all nodes, indicating that student activity is the dominant structural component of the educational ecosystem. Educational institutions, competencies, and learning objects are significantly less common, as the same educational resources are used repeatedly by many students. This distribution corresponds to the expected characteristics of large-scale educational interaction data.

Transforming educational events into polyadic interactions resulted in 4,861 hyperedges, as shown in Table 3.

Table 3

Hyperedge to context mapping

Context	Hyperedge
Assessment	1000
Twig	2972
BilimLand	889
Total	4861

Twig generated over 61% of all hyperlinks, demonstrating significantly richer patterns of educational interaction than other platforms. This observation indicates that Twig supports a much more diverse combination of learners, competencies, and learning objects, providing the greatest structural complexity among all the contexts studied.

Analysis of the educational contexts revealed significant differences in the number of unique participating objects, as shown in Table 4.

Table 4

Structural characteristics of platforms

Context	Hyperedges	Unique nodes used	Semantic role
Assessment	1000	667	Diagnostic effectiveness.
Twig	2972	1543	Engagement with digital content
BilimLand	889	682	Engagement with interactive resources

Twig contains more than twice as many unique educational objects as Assessment or BilimLand. Consequently, it demonstrates the greatest structural diversity and the most extensive educational interaction space within a single hypergraph. Intermediate result 1. A typified educational hypergraph structure was obtained, consisting of 2,509 vertices of five semantic categories and 4,861 hyperedges (Tables 2–4). These values serve as the initial data for calculating the structural invariance (SI) and adaptation cost (ACost) indices in Section 5. 2.

5. 2. Quantitative assessment of structural invariance and adaptation costs

After the construction of an invariant representation of educational data, a combined index based on the Jaccard coefficient is used to quantitatively assess the structural invariance between educational contexts

$$SI\left(C_i, C_j\right) = \frac{1}{2}\left(\frac{\left|V_i \cap V_j\right|}{\left|V_i \cup V_j\right|} + \frac{\left|E_i \cap E_j\right|}{\left|E_i \cup E_j\right|}\right), \quad (8)$$

where the first component characterizes the similarity of sets of educational entities (hypergraph nodes), and the second

characterizes the similarity of sets of educational interactions (hyperedges). Using the average value of two Jaccard coefficients allows us to simultaneously take into account the similarity of the composition of educational objects and the structure of their interactions when comparing different educational contexts.

Formula (8) quantitatively assesses the degree of structural similarity between two educational contexts through a joint analysis of the similarity of sets of educational entities and educational interactions. The index ranges from 0 to 1, where 1 corresponds to complete structural identity of the contexts under consideration, and 0 corresponds to the absence of common structural elements.

In addition to structural similarity, a transition or adaptation operator is introduced, characterizing the cost of transferring analytical procedures, integration rules, or educational services from one context to another. Let the expression

$$Ø_{ij} : H_i \rightsquigarrow H_j, \tag{9}$$

denotes the transition between contexts C_i and C_j. Then the corresponding adaptation cost is determined from the formula

$$ACost\left(C_i, C_j\right) = \lambda_1 \Delta_V\left(i,j\right) + \lambda_2 \Delta_E\left(i,j\right) + \lambda_3 \Delta_\Sigma\left(i,j\right), \tag{10}$$

provided

$$\lambda_1, \lambda_2, \lambda_3 \geq 0, \tag{11}$$

$\Delta_V(i, j)$ characterizes the differences between the sets of educational entities (hypergraph vertices) belonging to the contexts under consideration. $\Delta_E(i, j)$ characterizes the differences between the sets of hyperedges representing educational interactions in the contexts under consideration, and $\Delta_\Sigma(i, j)$ is designed to account for differences in the structural signatures of educational interactions. In the experiment under consideration, this component was set to zero since all educational contexts under consideration were converted to a single system of semantic signatures. The λ_1, λ_2 and λ_3 coefficients determine the relative importance of the corresponding components of the adaptation cost and can be normalized if necessary.

Increasing the minimum occurrence threshold sharply reduces the size of the vertex set, in accordance with Table 5.

Table 5

Characteristics of platform vertices and edges

Transition	Changing vertices ΔV	Changing hyperedges ΔE	*SI*_vertices	*SI*_edges	*SI*
Assessment → Twig	2056	3675	0.0000	0.0	0.0000
Twig → BilimLand	1425	3772	0.2118	0.0	0.1059
Assessment → BilimLand	1281	1800	0.0000	0.0	0.0000

The most noticeable reduction is observed between the first and second thresholds, indicating that many educational organizations appear only once in the dataset. After the third threshold is exceeded, the number of remaining nodes decreases much more slowly, suggesting that the underlying structure of the education system is relatively stable, with rarely encountered organizations contributing primarily to peripheral interactions. The transition between Twig and BilimLand demonstrates measurable structural similarities. Assessment has virtually no direct structural correspondence with either educational platform, as diagnostic testing follows a significantly different interaction model than content-based learning environments.

Let us consider the adaptation costs, which were calculated using equal weighting factors as shown in Table 6.

Table 6

Adaptation costs

Transition	Changes in vertices ΔV	Hyperedge changes ΔE	Changes to the scheme ΔS	ACost
Assessment → Twig	2056	3675	0	5731.0
Twig → BilimLand	1425	3772	0	5197.0
Assessment → BilimLand	1281	1800	0	3081.0

The highest adaptation costs are required for integrating Assessment with Twig, as the platforms under consideration have fundamentally different structures of educational interaction. Conversely, Assessment and BilimLand require the least structural transformations, despite their low structural similarity, indicating fewer differences in the number of educational objects and interaction signatures. Intermediate Result 2: Quantitative estimates of structural invariance *SI* (Table 5) and adaptation cost *ACost* (Table 6) were obtained for all pairs of educational contexts. These indicators serve as individual components of the integrated utility function *U*, calculated in Section 5. 5.

5. 3. Parametric sensitivity analysis of the model

The structural invariance index *SI* quantitatively characterizes the degree of similarity between educational contexts, the *ACost* function evaluates the structural complexity of their integration, and the integral utility function *U* combines the indicators of structural invariance, adaptation cost, and educational effectiveness. The combination of these indicators forms the analytical framework of the proposed invariant hypergraph model.

Flexibility, in this case, refers to the ability of the educational context to support a diversity of semantic entity types, as well as acceptable educational interactions. Furthermore, this indicator is considered a normalized characteristic of the structural potential of the educational context.

The parametric nature of the model is its key feature, and the simulation configuration can be defined by the following set, in accordance with the expression

$$\dot{E} = \left(\eta, \rho, p, \lambda_1, \lambda_2, \lambda_3, \alpha, \beta, \gamma\right), \tag{14}$$

where η is the minimum threshold for entity occurrences, ρ is the minimum hyperedge cardinality, and p is a perturbation parameter that controls the degradation or drift of the schema when transitioning between contexts. The perturbation parameter p is used to model the partial loss of correspondence between educational contexts. This also includes the absence of individual entities, incomplete mappings, incompatibility of local taxonomies, and differences in data representation schemes. Thus, the parametric nature of the model allows it to be used both for static comparison of observed systems and for studying the stability of structural invariants under controlled disturbances.

The proposed model implements a sequential process for analyzing multi-context educational data, comprising three interconnected stages:

1) semantic normalization of the source data by mapping platform-dependent entities into a unified space of educational types;

2) typified hypergraphs are formed, and structural invariance and adaptation cost indicators between educational contexts are calculated;

3) parametric analysis is performed to examine the impact of different integration scenarios and assess the comparative usefulness of educational platforms.

As a result of model execution, a set of quantitative characteristics *SI, ACost, U* is generated for each educational context and each pair of contexts. These characteristics are used to analyze structural compatibility, assess the complexity of integration, and comparatively rank educational platforms for decision-making.

Furthermore, this model allows for the separation of form invariance and semantic identity of educational data. In real-world educational systems, direct alignment of identifiers between platforms is often impossible due to the fact that competency catalogs, resource identifiers, assessment procedure codes, and user models are defined locally. In the proposed model, exact matching is treated as a special case rather than a mandatory prerequisite for integration. What should remain stable is not the coincidence of objects, but the permissible form of educational interactions and the interpretable distribution of their signatures. Thus, the proposed model considers invariant representation as a property of the semantic data normalization process, while structural invariance characterizes the degree of similarity of educational interactions between different contexts. This approach allows for comparison of educational contexts even with minimal overlap in source data.

The assessment has virtually no direct structural correspondence with any of the educational platforms, as diagnostic testing follows a significantly different interaction model than content-based learning environments, as illustrated in Fig. 1.

This sensitivity analysis reveals that increasing the perturbation magnitude *p* only slightly changes the invariance index for the "Twig-BilimLand" pair, while leaving the other transitions virtually unchanged. Thus, the identified structural similarities are due to the internal organization of educational interactions, rather than random coincidences.

We consider the sensitivity of the adaptation cost function ACost to changes in the weight coefficients λ. The experiment demonstrates how the adaptation cost changes with changes in the relative importance of the various components of the function, as shown in Fig. 2.

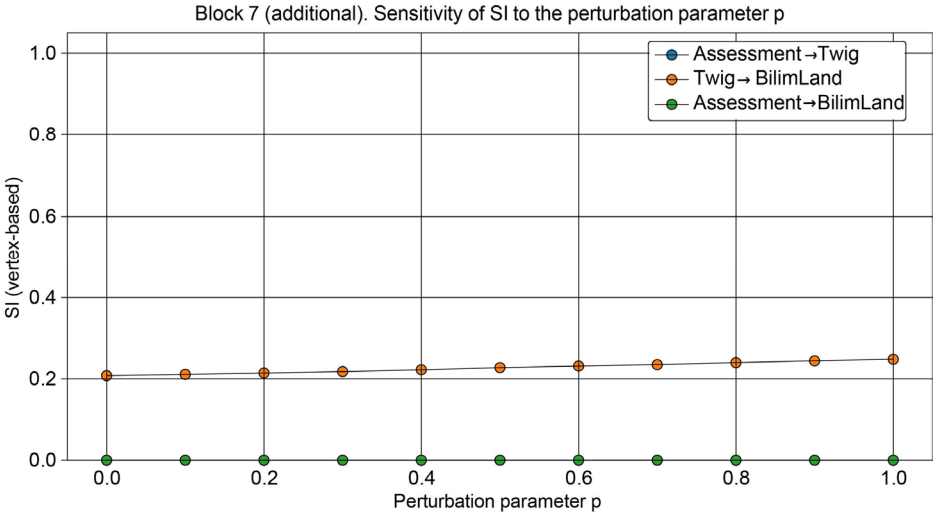


Fig. 1. Ratio of sensitivity to disturbance parameter

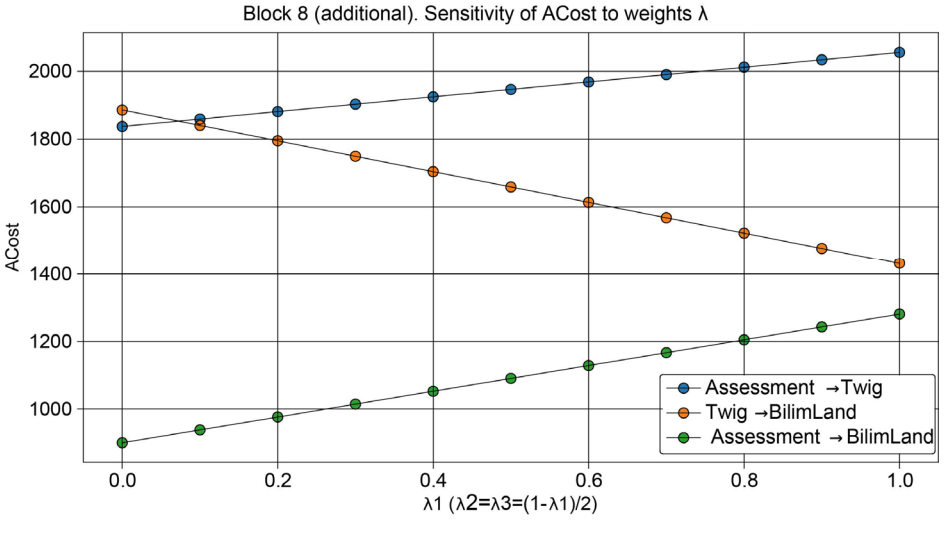


Fig. 2. Sensitivity of the adaptation cost function to changes in weight coefficients

The analysis reveals that increasing the weight assigned to vertex differences gradually increases the overall adaptation cost, while emphasizing hyperedge similarity reduces the overall cost. This confirms that the proposed model allows for flexible prioritization of different integration objectives depending on the practical implementation requirements. Intermediate Result 3. Sensitivity analysis demonstrated the stability of the *SI* and *ACost* metrics across the studied ranges of variation of the perturbation parameter p and weighting coefficients λ (Fig. 1, 2). Within the considered scenarios, changing the parameters did not change the comparative ranking of the platforms, demonstrating the robustness of the results of the baseline computational experiment.

5. 4. Results of the computational experiment

The invariant educational hypergraph model was evaluated using empirical data collected from three independent digital educational environments: Assessment, Twig, and BilimLand. Each platform represents a distinct educational context characterized by different interaction scenarios, data structures, and semantic organization.

The experimental process included:

– construction of a unified, typed educational hypergraph;

– identification of educational entities and hyperedges;

– extraction of context-sensitive subgraphs;

– assessment of structural invariance under simulated disturbances;

– assessment of adaptation costs between platforms;

– calculation of an integrated utility function for comparative evaluation.

The following parameters given in Table 7 were used in the baseline experiment.

Table 7

Parameters of the hypergraph model

No.	Parameter	Value
0	min_entity_count	1.000000
1	min_hyperedge_size	2.000000
2	perturbation_p	0.100000
3	random_seed	42.000000
4	lambda_1	1.000000
5	lambda_2	1.000000
6	lambda_3	1.000000
7	alpha	0.333333
8	beta	0.333333
9	gamma	0.333333

These settings ensured that all detected educational entities were included in the analysis, while providing an objective comparison of structural similarity, adaptation complexity, and educational effectiveness.

Unlike traditional graph analyses, which only examine pairwise relationships, the proposed model preserves complete educational events as polyadic interactions. Consequently, the resulting quantitative metrics characterize not only the volume of educational data but also the structural organization of interactions on heterogeneous digital learning platforms.

5. 5. Assessing the comparative effectiveness of educational platforms

For comparative analysis of contexts, a global utility function is introduced. Let the integral utility of an educational context be defined by the expression

$$U_{CTX} = \alpha F + \beta\left(1 - \overline{ACost}\right) + \gamma \overline{LR}, \quad (12)$$

where F is the normalized flexibility of the context, $\overline{ACost}$ is the normalized cost of adaptation, $\overline{LR}$ is the normalized indicator of the educational outcome, and α, β, γ are the preference coefficients satisfying the conditions

$$\alpha, \beta, \gamma \geq 0,\ \alpha + \beta + \gamma = 1. \quad (13)$$

The final result summarizes the intermediate results of Sections 5.1–5.3 (hypergraph structural characteristics, *SI*, *ACost*) into a single integrated assessment of educational platforms.

The integrated utility function combines structural flexibility, adaptation complexity, and educational effectiveness into a single evaluation criterion, as shown in Table 8.

Table 8

Comparison of educational platforms

Context	F	AC	LR	U
Assessment	0.8	0.248526	0.697428	0.581985
Twig	0.8	0.634794	1.000000	0.811598
BilimLand	0.8	0.267060	0.691899	0.586320

Here, the Twig platform demonstrates the highest utility value due to its combination of extensive educational interactions, the largest number of unique objects, and the most active educational activities. Although adaptation costs remain relatively high, this disadvantage is offset by a significantly richer educational structure.

The final integrated assessment of educational platforms is represented in accordance with Fig. 3.

Fig. 3 illustrates the values of the integral utility function components for the studied educational platforms after normalization of the indicators. Structural flexibility F takes the same value for all contexts, indicating a comparable level of diversity in the semantic types used. The main differences are formed by the educational performance components LR and the normalized adaptation cost (1 – *ACost*). The highest value for both indicators is observed for the Twig platform, resulting in the maximum integral utility function value $U = 0.81$. Assessment and BilimLand demonstrate similar values for the final utility (around 0.59), despite differences in individual components. Thus, the integral function enables a multi-criteria comparison of educational platforms, simultaneously considering their structural characteristics, integration complexity, and educational performance.

To summarize our results and identify the distinctive characteristics of the proposed model, it is compared with known approaches to the integration and analysis of heterogeneous data. The comparison was based on the type of modeled relationships, the presence of a quantitative assessment of structural compatibility, the degree of automation of semantic comparison, and the nature of experimental validation. The results of the comparative analysis are given in Table 9.

A comparative analysis reveals that existing approaches address individual components of the problem under consideration: modeling high-order relationships, processing heterogeneous graphs, and matching data schemas. The IHIA model combines a polyadic hypergraph representation of educational interactions with a quantitative assessment of structural compatibility based on the *SI* and *ACost* metrics. Additionally, the model provides an integrated assessment of the usefulness of *U* and experimental validation using data from several independent educational platforms. Furthermore, semantic matching of entities in the current implementation is performed based on predefined normalization rules, which should be considered a limitation of the proposed approach.

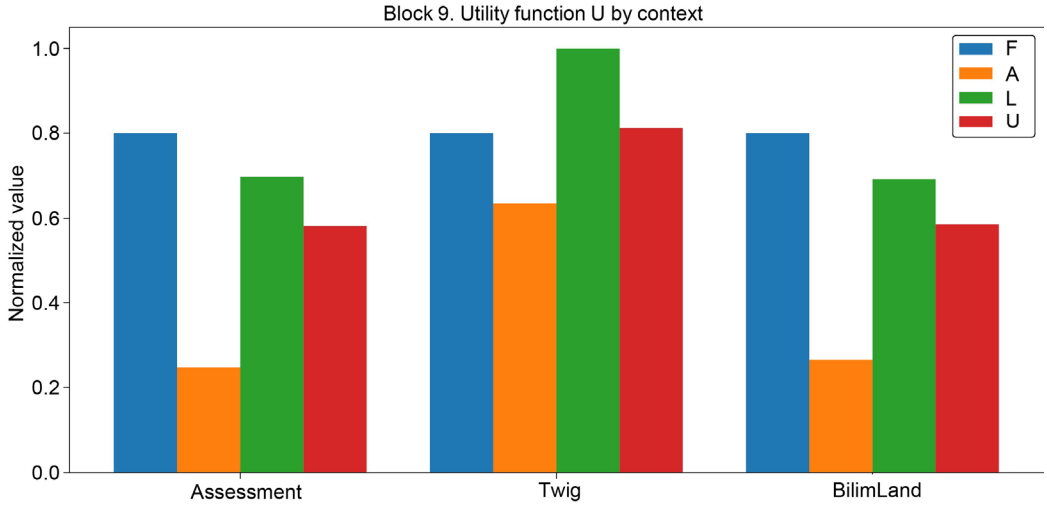


Fig. 3. Integrated assessment of educational platforms

Table 9

Comparing the proposed approach (IHIA) with known approaches for integrating heterogeneous data

Approach	Connection type	Quantitative compatibility assessment	Automated semantic matching	Validation on independent educational platforms
Hypergraph Neural Networks [2]	Polyadic (hyperedges)	Not provided (one hypergraph)	–	Not conducted (single data source)
Heterogeneous Graph Transformer, HGT [12]	Binary (typed graph)	Not provided	Requires a pre-defined, unified graph schema	Within a single academic network
Ontology Matching [13, 14]	— (scheme matching)	Not provided	Partial (structural/lexical methods)	Heterogeneous general-purpose ontologies, not educational platforms
EduCross [11]	Bipartite hypergraph	Not provided	—	Within a single collection of multimodal materials
The proposed approach (IHIA)	Polyadic (hyperedges, formula (5))	SI, ACost, U (formulaes (8)–(13); Tables 5, 6, 8)	Partial, based on pre-defined normalization rules (section 4.3)	Three independent platforms: Assessment, Twig, BilimLand (Tables 1–4)

6. Discussion of the structural compatibility of educational platforms and the practical applicability of the model for integrating analytics and invariant graphs

The results of our study demonstrate that the proposed Invariant Hypergraph Integration and Analytics (IHIA) model could be used as a means of unifying educational data and as a formal basis for analyzing multi-context educational ecosystems. While traditional graph models represent educational events as a set of binary relationships, the typed hypergraph preserves the polyadic nature of educational interactions by uniting all participants in the relevant educational process within a single hyperedge. The proposed approach was validated in a computational experiment in which data from the Assessment, Twig, and BilimLand platforms were converted to a single invariant representation without changing the mathematical structure of the model (Tables 1–4, Fig. 3). The experiment demonstrated that the presence of a common data representation does not necessarily imply high structural compatibility. Quantitative analysis (Table 6, Fig. 3) allowed us to identify differences in the organization of educational interactions that cannot be detected using only traditional data integration schemes.

Our results are consistent with modern studies demonstrating the advantages of hypergraphs in modeling high-order relations and complex interaction structures [21]. However, unlike studies where the hypergraph is used to solve a separate applied task of annotating educational images [22], predicting student performance [23], or tracking learners' knowledge [24], in the present work the hypergraph is used in a different role: as a universal analytical tool for the integration, quantitative comparison, and assessment of the structural compatibility of educational platforms, rather than for solving a single applied task. Unlike [2, 12], in which the structural compatibility of multiple data sources is not assessed quantitatively, the proposed model introduces explicit metrics SI and ACost for this purpose (formulaes (8), (10); Tables 5, 6), which allows for an objective comparison of platforms, and not only descriptive, as in [11].

Despite belonging to the same subject area, the educational platforms under study are characterized by different structures of educational interactions (Tables 4, 8). These differences are due to the functional purpose of the platforms: Assessment is focused on diagnostic assessment, while Twig and BilimLand support longer-term interactions between students and digital educational resources. The functional purpose of the educational platforms leads to the formation of a more complex structure of educational connections (Table 3). This interpretation is consistent with modern research in the field of learning analytics, where various digital platforms are considered as complementary sources of information on the educational process [8, 15, 16]. Unlike existing studies, which use structural heterogeneity primarily to

improve the quality of predictive models, this paper examines structural heterogeneity for the first time as an independent object of quantitative research.

Perhaps one of the most important findings is that even with a single hypergraph representation, structural invariance remains relatively low (Fig. 1). This indicates that the formal compatibility of data structures does not automatically imply the semantic compatibility of educational interactions. In most studies on hypergraph-based neural networks, the concept of invariance is used to mean the mathematical stability of graph representations [21]. In contrast, in this paper, invariance is considered as a quantitative measure of the compatibility of educational interactions across contexts: low SI values do not indicate the inability of the model but rather show that the model distinguishes patterns of educational interactions. Another methodological contribution of the work is the adaptation cost indicator ACost (Table 6, Fig. 2): the results indicate that the complexity of integration depends primarily on differences in the structures of educational interactions, rather than on differences in individual educational entities.

In practice, the methodology facilitates the assessment of platform interactions, evaluation of digital ecosystem architecture, integration planning, and the selection of platforms for implementation. Twig's consistent superiority throughout the sensitivity analysis (Fig. 1, 2, Table 7) indicates that the ranking reflects structural properties rather than parametric artifacts. Unlike studies aimed at optimizing predictive metrics, the proposed framework facilitates strategic decision-making regarding platform interactions.

Compared to existing approaches, the proposed solution has two main limitations. First, unlike ontological matching methods [13, 14], which provide partial automation of schema matching, the semantic matching of entities between educational platforms in this model is performed based on predefined normalization rules rather than being automatically learned. Second, unlike studies validated on broad open benchmarks [2, 12], our experimental verification is limited to three platforms within a single Bilim Group ecosystem (Table 1), limiting the generalizability of the findings to additional LMS and MOOC platforms.

Further advancement of our research involves automating the semantic matching of educational entities using ontology and knowledge graph matching methods. A separate study area is the expansion of the experimental validation of the model to independent LMS and MOOC platforms with different data representation schemes. This will allow us to assess the model's portability across educational ecosystems and determine the limits of its applicability. Research into methods for automatically adjusting the model's weighting coefficients based on the characteristics of the analyzed educational contexts also holds promise.

7. Conclusions

1. A multi-context educational environment has been formalized as a typed invariant hypergraph that preserves the polyadic nature of interactions (unlike binary graph models) regardless of the platform structure. This was confirmed by constructing a hypergraph consisting of 2,509 nodes and 4,861 hyperedges based on 3,141 records from three educational platforms.

2. We have obtained quantitative estimates for the structural compatibility and adaptation costs of three pairs of educational contexts. For Twig-BilimLand, the *SI* index was 0.1059, while for the remaining pairs, *SI* was 0. The *ACost* values were 5,731 for Assessment-Twig, 5,197 for Twig-BilimLand, and 3,081 for Assessment-BilimLand.

3. A parametric model design has been developed, including the perturbation parameter p, thresholds η and ρ, as well as weighting coefficients λ_1–λ_3 and α–γ. A sensitivity analysis revealed that the comparative ranking of the platforms was maintained when varying the studied parameters. For the Twig-BilimLand pair, the base SI value was 0.1059, while for the other two pairs it remained equal to 0.

4. A computational experiment was conducted based on data from three platforms ("Assessment," "Twig," and "BilimLand"). The "Twig" platform achieved the highest overall utility value ($U = 0.81$) due to its high educational effectiveness and low adaptation costs, while the "Assessment" and "BilimLand" platforms demonstrated similar but lower values (around 0.59).

5. A parametric model design has been developed, including the perturbation parameter p, thresholds η and ρ, as well as weighting coefficients λ_1–λ_3 and α–γ. A sensitivity analysis revealed that the comparative ranking of the platforms was maintained when varying the studied parameters. For the Twig–BilimLand pair, the baseline SI value was 0.1059, while for the other two pairs it remained equal to 0.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study and the results reported in this paper.

Funding

This study was/is funded by the Science Committee of the Ministry of Science and Higher Education, the Republic of Kazakhstan (grant no. AP23488869).

Data availability

The manuscript has associated data in the data warehouse.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

Author contributions

Kainizhamal Iklassova: Conceptualization; Project administration; Writing – original draft; Writing – review & editing; **Aigul Shaikhanova:** Methodology; Writing – orig-

inal draft; **Gulnar Kim** Writing – original draft; Software; Validation; **Kanat Myrzakhanov:** Resources, Data Curation; Validation; **Abduraimova Bayan** Formal analysis**, Madina Bazarova** Investigation; Validation.

All authors have read and agreed to the published version of the article and are jointly responsible for the reliability, accuracy, and integrity of the data and research results presented in the article.

References

1. Wang, Y., Wang, Z., Yu, X., Wang, X., Song, J., Yu, D.-J., Ge, F. (2024). MORE: a multi-omics data-driven hypergraph integration network for biomedical data classification and biomarker identification. Briefings in Bioinformatics, 26 (1). https://doi.org/10.1093/bib/bbae658
2. Feng, Y., You, H., Zhang, Z., Ji, R., Gao, Y. (2019). Hypergraph Neural Networks. Proceedings of the AAAI Conference on Artificial Intelligence, 33 (01), 3558–3565. https://doi.org/10.1609/aaai.v33i01.33013558
3. Antelmi, A., Cordasco, G., Polato, M., Scarano, V., Spagnuolo, C., Yang, D. (2023). A Survey on Hypergraph Representation Learning. ACM Computing Surveys, 56 (1), 1–38. https://doi.org/10.1145/3605776
4. Wang, X., Bo, D., Shi, C., Fan, S., Ye, Y., Yu, P. S. (2023). A Survey on Heterogeneous Graph Embedding: Methods, Techniques, Applications and Sources. IEEE Transactions on Big Data, 9 (2), 415–436. https://doi.org/10.1109/tbdata.2022.3177455
5. Xia, F., Sun, K., Yu, S., Aziz, A., Wan, L., Pan, S., Liu, H. (2021). Graph Learning: A Survey. IEEE Transactions on Artificial Intelligence, 2 (2), 109–127. https://doi.org/10.1109/tai.2021.3076021
6. Ji, S., Pan, S., Cambria, E., Marttinen, P., Yu, P. S. (2022). A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. IEEE Transactions on Neural Networks and Learning Systems, 33 (2), 494–514. https://doi.org/10.1109/tnnls.2021.3070843
7. Lang, C., Siemens, G., Wise, A., Gasevic, D. (Eds.) (2017). Handbook of Learning Analytics. Solaresearch. https://doi.org/10.18608/hla17
8. OECD Digital Education Outlook 2023 (2023). OECD Digital Education Outlook. OECD Publishing. https://doi.org/10.1787/c74f03de-en
9. Xie, C., Zhong, W., Xu, W., Mueller, K. (2018). Visual Analytics of Heterogeneous Data Using Hypergraph Learning. ACM Transactions on Intelligent Systems and Technology, 10 (1), 1–26. https://doi.org/10.1145/3200765
10. Zhou, D., Huang, J., Schölkopf, B. (2007). Learning with Hypergraphs: Clustering, Classification, and Embedding. Advances in Neural Information Processing Systems 19, 1601–1608. https://doi.org/10.7551/mitpress/7503.003.0205
11. Li, M., Zhou, S., Chen, Y., Huang, C., Jiang, Y. (2024). EduCross: Dual adversarial bipartite hypergraph learning for cross-modal retrieval in multimodal educational slides. Information Fusion, 109, 102428. https://doi.org/10.1016/j.inffus.2024.102428
12. Hu, Z., Dong, Y., Wang, K., Sun, Y. (2020). Heterogeneous Graph Transformer. Proceedings of the Web Conference 2020, 2704–2710. https://doi.org/10.1145/3366423.3380027
13. Euzenat, J., Shvaiko, P. (2013). Ontology Matching. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38721-0
14. Paulheim, H. (2016). Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web, 8 (3), 489–508. https://doi.org/10.3233/sw-160218
15. Siemens, G., Long, P. (2014). Penetrating the Fog: Analytics in Learning and Education. Italian Journal of Educational Technology, 22 (3). https://doi.org/10.17471/2499-4324/195
16. Romero, C., Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. WIREs Data Mining and Knowledge Discovery, 10 (3). https://doi.org/10.1002/widm.1355
17. Berge, C. (1989). Hypergraphs: combinatorics of finite sets. North Holland: Distributors for the U.S.A. and Canada, Elsevier Science Pub. Co. Available at: http://archive.org/details/hypergraphscombi0000berg
18. Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. D., Gutierrez, C. et al. (2021). Knowledge Graphs. ACM Computing Surveys, 54 (4), 1–37. https://doi.org/10.1145/3447772
19. Fatemi, B., Taslakian, P., Vazquez, D., Poole, D. (2020). Knowledge Hypergraphs: Prediction Beyond Binary Relations. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, 2191–2197. https://doi.org/10.24963/ijcai.2020/303
20. Huang, J., Yang, J. (2021). UniGNN: a Unified Framework for Graph and Hypergraph Neural Networks. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 2563–2569. https://doi.org/10.24963/ijcai.2021/353
21. Kim, S., Lee, S. Y., Gao, Y., Antelmi, A., Polato, M., Shin, K. (2024). A Survey on Hypergraph Neural Networks: An In-Depth and Step-By-Step Guide. Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 6534–6544. https://doi.org/10.1145/3637528.3671457
22. Li, H., Wang, J., Du, X., Hu, Z., Yang, S. (2023). KBHN: A knowledge-aware bi-hypergraph network based on visual-knowledge features fusion for teaching image annotation. Information Processing & Management, 60 (1), 103106. https://doi.org/10.1016/j.ipm.2022.103106
23. Li, M., Cheng, Y., Bai, L., Cao, F., Lv, K., Liang, J., Lio, P. (2025). EduLLM: Leveraging Large Language Models and Framelet-Based Signed Hypergraph Neural Networks for Student Performance Prediction. Proceedings of the 42nd International Conference on Machine Learning. Available at: https://proceedings.mlr.press/v267/li25e.html
24. Ma, F., Zhu, C., Lei, P. (2025). Hypergraph-Driven High-Order Knowledge Tracing with a Dual-Gated Dynamic Mechanism. Applied Sciences, 15 (15), 8617. https://doi.org/10.3390/app15158617