

Inferring Prerequisite Knowledge Concepts in Educational Knowledge Graphs: A Multi-criteria Approach

Rawaa Alatrash¹[0000–0003–2192–028X], Mohamed Amine
Chatti¹[0000–0002–1311–7852], Nasha Wibowo¹, and Qurat Ul
Ain¹[0000–0003–2691–0267]

Social Computing Group, Faculty of Computer Science, University of
Duisburg-Essen, Duisburg, Germany

Abstract. Educational Knowledge Graphs (EduKGs) organize various learning entities and their relationships to support structured and adaptive learning. Prerequisite relationships (PRs) are critical in EduKGs for defining the logical order in which concepts should be learned. However, the current EduKG in the MOOC platform CourseMapper lacks explicit PR links, and manually annotating them is time-consuming and inconsistent. To address this, we propose an unsupervised method for automatically inferring concept PRs without relying on labeled data. We define ten criteria based on document-based, Wikipedia hyperlink-based, graph-based, and text-based features, and combine them using a voting algorithm to robustly capture PRs in educational content. Experiments on benchmark datasets show that our approach achieves higher precision than existing methods while maintaining scalability and adaptability, thus providing reliable support for sequence-aware learning in CourseMapper.

Keywords: Educational Knowledge Graphs · MOOCs · Prerequisite concept identification · Relation extraction · Knowledge structure

1 Introduction

Educational Knowledge Graphs (EduKGs) play a vital role in organizing and representing key entities and their relationships in digital learning environments, supporting knowledge discovery, concept navigation, and learner modeling [1]. In our MOOC platform CourseMapper, EduKGs provide learners with a structured overview of the main knowledge concepts covered in learning materials to enhance understanding and engagement [2,3,1]. However, a major limitation of current EduKG structures in CourseMapper is the lack of prerequisite relationships (PRs), which define concept dependencies where understanding one concept is necessary before learning another. These relationships determine the logical sequence for mastering topics and support personalized sequence recommendations [15]. Without explicit PRs, learners may struggle to identify which concepts should be mastered first, which can hinder self-directed learning and

adaptive recommendation systems. Prior research has explored end-to-end learning, feature-based, and hybrid methods for PR extraction [6,15,12,5]. End-to-end learning-based methods depend on domain-specific labeled data and static features, which limits their scalability. Feature-based methods remain widely used and rely on graph structures, text content, or Wikipedia data to capture concept dependencies [12,7]. However, these methods have several limitations. They often overlook how semantic and structural information interact and they often rely only on concept order. Moreover, most feature-based approaches assess each criterion individually instead of combining their contributions for effective PR identification. To overcome these limitations, we propose an unsupervised, multi-criteria approach that combines the contributions of document-based, Wikipedia hyperlink-based, graph-based, and text-based features using a voting algorithm, aiming to effectively extract PRs and enrich the structure of EduKGs in CourseMapper.

2 Background and Related Work

2.1 EduKG Generation in CourseMapper

Educational Knowledge Graphs (EduKGs) represent diverse entities and their relationships within a learning context, such as concepts, instructors, courses, and institutions [1]. In our MOOC platform CourseMapper [2], instructors can upload Learning Materials (LM), for which an EduKG is automatically generated. This includes entities such as LM, Slides (S), Main Concepts (MC), Related Concepts (RC), Categories (Cat), and Learners (L), linked by relationships such as CONTAINS, CONSISTS_OF, RELATED_TO, HAS_CATEGORY, HAS_READ, UNDERSTOOD (U), and DID_NOT_UNDERSTAND (DNU). MCs are identified in each slide using a keyphrase extraction algorithm. These MCs undergo an expansion process via DBpedia Spotlight, which adds additional knowledge concepts as RCs and Cats. The RCs and Cats are then filtered to retain only the most relevant ones using cosine similarity in the EduKG[3]. This structure provides learners with an organized view of key concepts in the LM. The EduKG can be further enhanced by adding prerequisite relationships, which model the learning sequence among concepts and support adaptive learning pathways [15]. Currently, CourseMapper’s EduKG does not include PRs, limiting its support for sequence-aware learning.

2.2 Prerequisite Relationship Extraction

PRs describe how concepts depend on each other in learning pathways. Methods for identifying PRs include end-to-end learning-based, feature-based, and hybrid approaches. End-to-end learning-based methods rely on deep learning to detect PRs from large datasets [6,19,17]. Feature-based methods use manually designed features [15,12,7] such as graph-based features [12], which capture structural characteristics of graphs; text-based features, which use descriptions and content [23,8]; and Wikipedia-based features, which exploit hyperlinks,

clickstreams, and categories [18,5,16]. Hybrid methods combine both end-to-end learning-based and feature-based methods [12,5]. Many existing end-to-end learning-based methods for PR inference depend on domain-specific labeled data and simplistic features, limiting scalability and adaptability. Manual annotation is time-consuming and inconsistent. Most feature-based methods consider criteria independently, often ignoring the interaction between semantic and structural information. To overcome these issues, we propose an unsupervised multi-criteria approach combining ten distinct criteria document-based, Wikipedia hyperlink-based, graph-based, and text-based features. These criteria integrate semantic, structural, and contextual information through a **voting algorithm** that balances their contributions and mitigates noise from any single criterion. Our method evaluates all concept pairs in the EduKG, which is crucial for unstructured content where structural criteria alone may be misleading. Therefore, incorporating additional criteria that leverage external resources and contextual information can enhance the accuracy of PR inference. Our approach, tested on two public datasets, achieves a precision of 1.0, demonstrating robustness, domain independence, and effective PR extraction.

3 Methodology

To define the likelihood of one concept being a prerequisite of another, we developed and adopted a set of criteria, derived from the literature review and adapted for our context. These criteria are categorized into document-based features, Wikipedia hyperlink-based features, graph-based features, hierarchy-based features, and text-based features. Our multi-criteria approach begins by examining concept pairs against these criteria, each offering a unique perspective. Their outputs are then aggregated using a voting algorithm, followed by a thresholding step to infer whether a PR exists. These criteria are applied to all concept pairs in EduKG in CourseMapper and examined in both directions to ensure comprehensive identification. This study excludes concept synonyms and automatic redirections, as these are managed during the EduKG construction on the platform CourseMapper, ensuring only unique nodes without duplicates are extracted [3]. Before outlining the model structure, we adopt three core assumptions: **Transitivity** [5,4], where if $A \xrightarrow{\text{PR}} B$ and $B \xrightarrow{\text{PR}} C$, then $A \xrightarrow{\text{PR}} C$; **Asymmetry** [5,4], where if $B \rightarrow A$, then $A \nrightarrow B$; and **No Self-Connection** [4], a concept cannot be a prerequisite of itself. Table 1 lists each criterion with its name, abbreviation, and description.

3.1 Prerequisite Relationships Extraction Features:

Document-based Features: This category depends on the document structure to obtain the **temporal order** (TemO) of concepts within an article where the concept introduced first serves as a prerequisite for the subsequent [19]. For instance, in discussing *Clustering*, the term *machine learning* is likely introduced before *Clustering* as understanding the former is essential for grasping the latter.

Table 1: Categories of Criteria with Full Names, Abbreviations, and Descriptions.

Category	Full name	Abbreviations	Description
Document based features	Temporal order	TemO	c_1 precedes c_2 in the learning material or channel
Wikipedia Hyperlinks based features	Hyperlinks on each other’s articles	HL-A	c_1 hyperlink is mentioned within c_2 ’s article
	Hyperlinks on each other’s abstract	HL-Ab	c_1 hyperlink appears in the abstract of c_2
	Hyperlinks on each other’s related concepts’ abstracts	HL-RCA	c_1 hyperlink is given on c_2 ’s related concept’s abstract
	Reference distance	RefD	c_1 hyperlink is mentioned in the related concepts of c_2
	Inbound links and outbound links ratio	IOLR	c_1 has higher ratio of the number of in/outbound links pointing to c_2
Graph based features	One concept is a category of another concept	CatCon	c_1 is a category of c_2 in DBPedia hierarchy
	One concept is the direct super-category of another concept	SuperCatCon	c_1 is a super category of c_2 in DBPedia hierarchy
Hierarchy based features	CourseMapper Hierarchy	CMH	first occurrence of c_1 in $channel_1$ precedes first occurrence of c_2 in $channel_2$
Text-based features	BERTopic and Shannon’s Entropy	BERTropy	c_1 has higher topic entropy than c_2

$TemO(c_1, c_2)$ is assigned the value 1 if the position $P(c_1)$ of concept c_1 appears before the position $P(c_2)$ of concept c_2 ; otherwise, 0.

$$TemO(c_1, c_2) = \begin{cases} 1, & \text{if } P(c_1) < P(c_2) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

TemO applies only to MCs, as it relies on their order of appearance in the LM; related concepts are excluded because they are not explicitly mentioned.

Wikipedia Hyperlinks-based Features: Wikipedia has been exploited to find potential PRs among concepts by analyzing the hyperlink structure of their corresponding Wikipedia articles [20,5].

Hyperlinks on each other’s articles (HL-A): In this criteria, we examine whether the hyperlinks for each concept appear in the corresponding Wikipedia articles of other concepts. For example, if the hyperlink for the concept *Artificial Intelligence* (c_1) is found in the Wikipedia article for *Machine Learning* (c_2), it suggests that *Artificial Intelligence* serves as a prerequisite for understanding

Machine Learning. This implies that c_1 is a prerequisite to c_2 . These relationships are formalized as follows:

$$HL - A(c_1, c_2) = \begin{cases} 1, & \text{if } c_1 \in H_{c_2} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $HL - A(c_1, c_2)$ determines whether the hyperlink for concept c_1 exists in the Wikipedia article of concept c_2 , where in this case 1 is assigned, otherwise 0. H_{c_2} represents the set of concepts that are hyperlinked within the Wikipedia article for concept c_2 . Sometimes, hyperlinks in Wikipedia articles do not indicate a PR especially in long articles or hyperlinks at the end of the article. For example, if *Supervised Learning* appears at the end of the article *Machine Learning* it could misleadingly suggest that *Supervised Learning* is a prerequisite to *Machine Learning*. Therefore, we need to consider criteria focused on the abstracts of articles for such special cases [10].

Hyperlinks on each other's abstract (HL-Ab): This criterion is to find out if c_1 hyperlinks are included in c_2 abstract. The abstract typically provides a brief definition, and if hyperlink for concept c_1 is mentioned in the abstract of concept c_2 , it suggests that c_1 is likely a prerequisite for c_2 [10]. For instance, if *Machine Learning* is referenced in the abstract of *Supervised Learning* this implies that understanding *Machine Learning* is necessary before learning *Supervised Learning* [10]. The relationships are formulated as follows:

$$HL - Ab(c_1, c_2) = \begin{cases} 1, & \text{if } c_1 \in AH_{c_2} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

$HL - Ab(c_1, c_2)$ indicates whether c_1 is a prerequisite to c_2 , where 1 is assigned in case hyperlink for concept c_1 belongs to AH_{c_2} which is the set of concepts hyperlinked within the abstract of the Wikipedia article for concept c_2 , otherwise 0.

Hyperlinks on each other's related concepts' abstracts (HL-RCA): Inspired by the idea of analyzing the importance of related concepts to infer PRs [9] and the role of hyperlinks in abstracts, we developed this criterion. This involves analyzing the abstracts of related concepts for each concept pair. For instance, consider the concepts *artificial intelligence* and *machine learning*. A related concept of *machine learning*, such as *artificial neural network*, is likely to reference a hyperlink for concept *artificial intelligence* in the abstract of its Wikipedia article. In contrast, a concept related to *artificial intelligence*, such as *knowledge representation*, does not mention the hyperlink of *machine learning* concept within the abstract of its article. This pattern arises because related concepts often share common prerequisite concepts. In this example, *artificial intelligence* serves as a prerequisite for both *machine learning* and *artificial neural network* as given in the following formula:

$$HL - RCA(c_1, c_2) = \begin{cases} 1, & \text{if } c_1 \in A_{RC_{c_2}} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where $A_{RC_{c_2}}$ is the set of concepts that are hyperlinked within the abstract of the Wikipedia article of the related concepts of c_2 . $HL - RCA(c_1, c_2)$ will be 1 in case c_1 hyperlinks are present in $A_{RC_{c_2}}$, otherwise 0.

Reference distance (RefD): It is defined as the reference distance between concept pair [9]. This measures the proximity of Wikipedia articles based on their reference links and can be an excellent indicator for finding PRs. The idea here is that for each concept pair, we are comparing the asymmetry of c_1 's related concept mentioning c_2 and c_2 's related concepts mentioning c_1 . For example, most of *Data Mining* related concepts mention *Data Science* in their articles. However, none of *Data Science* related concepts mention *Data Mining* in their articles. We adapted this equation to fit our system as follows:

$$RefD_{(c_1, c_2)} = \frac{\sum_{i=1}^k r(R_{c_i}, c_2) w(R_{c_i}, c_1)}{\sum_{i=1}^k w(R_{c_i}, c_1)} - \frac{\sum_{i=1}^k r(R_{c_i}, c_1) w(R_{c_i}, c_2)}{\sum_{i=1}^k w(R_{c_i}, c_2)} \quad (5)$$

where $C = \{c_1, \dots, c_k\}$ is the concept space, R_{c_i} is a set of related concepts to c_i , $w(R_{c_i}, c_1)$ is cosine similarity between R_{c_i} and c_1 , $r(R_{c_i}, c_1)$ the relationship between related concept of c_i and c_1 , where R_{c_i} refers to c_1 , meaning whether the related concept of c_i includes c_1 . The outcome of RefD is then compared against a threshold to determine whether a PR is indicated as shown in the equation 6 below.

$$RefD(c_1, c_2) = \begin{cases} 1, & \text{if } -1 < RefD(c_1, c_2) < -\theta \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

c_1 is considered a prerequisite of c_2 if the RefD result falls between -1 and $-\theta$. Conversely, c_2 is assumed to be a prerequisite of c_1 if the RefD result is between θ and 1. While between $-\theta$ and θ indicates no prerequisite can be inferred, where θ is empirically determined. This is defined as the following:

$$RefD(c_2, c_1) = \begin{cases} 1, & \text{if } \theta < RefD(c_1, c_2) < 1 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

Inbound links and outbound links ratio (IOLR): This criterion evaluates a concept's connectivity by comparing its inbound and outbound links [10,5]. Foundational concepts usually receive more inbound links, while advanced ones tend to have more outbound links. To reduce bias from generally popular concepts, we compute the inbound-to-outbound link ratio (IOL), where a higher IOL indicates that a concept is more foundational and likely to be a prerequisite, as given below:

$$IOL(c_i) = \frac{\sum inboundlinks}{\sum outboundlinks} \quad (8)$$

Then, we define the ratio to determine the direction of the relationship, as indicated by the equation $IOLR(c_1, c_2)$.

$$IOLR(c_1, c_2) = \begin{cases} 1, & \text{if } IOL(c_1) > IOL(c_2) \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

Graph-based Features: This category adopts the hierarchical structure of DBpedia, where the categories in DBpedia represent the hierarchical relationships between concepts.

One concept is a category of another concept (CatCon): Categories outline the essential topics of concepts, and understanding a concept often requires knowledge of its foundational elements defined by its category [20,5]. Based on the DBpedia hierarchy, if c_2 is classified under c_1 , we can infer that c_1 is a category of c_2 and thus a prerequisite for c_2 . To formulate this, we define K_{c_2} as the category associated with c_2 and vice versa in case of K_{c_1} . This is given in the equation below:

$$CatCon(c_1, c_2) = \begin{cases} 1, & \text{if } c_1 \in K_{c_2} \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

$CatCon(c_1, c_2)$ indicates that c_1 belongs to the category associated with c_2 . In other words, c_2 is categorized under c_1 . This, strongly implies that c_1 serves as a prerequisite for c_2 .

One concept is the direct super-category of another concept (SuperCatCon): Another useful signal for identifying PRs is the direct super-category of a concept, defined as the category of its category in DBpedia [20]. This follows a transitive logic: if c_1 (*Machine Learning*) is a category of c_2 (*Unsupervised Learning*), and c_2 is a category of c_3 (*K-Means*). Thus, c_1 is the super-category of c_3 , which means c_1 is inferred as a prerequisite of c_3 . Let $K_{c_1}^{up}$ denote the direct super-category of c_1 , PR exists when these intersect with the categories of c_2 .

$$SuperCat(c_1, c_2) = |K_{c_1}^{up} \cap K_{c_2}| \quad (11)$$

This is also repeated to calculate the existence of prerequisite in both directions $super(c_2, c_1)$. Then, we compare their asymmetry to avoid intertwining categories. We want to prevent scenarios where some super-categories of c_2 are also categories of c_1 for $SuperCat(c_1, c_2)$ or for $SuperCat(c_2, c_1)$, which would imply mutual prerequisites, as follows:

$$SuperCatCon(c_1, c_2) = \begin{cases} 1, & \text{if } len(SuperCat(c_1, c_2)) > len(SuperCat(c_2, c_1)) \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

When $SuperCatCon(c_1, c_2)$ is 1, indicates that more of c_1 's super-categories are also categories of c_2 suggesting that c_1 is a prerequisite of c_2 .

Hierarchy-based Features: This category captures the structural organization of educational content, reflecting concept relationships based on their positions within hierarchical formats. This approach can be applied to all educational platforms that follow a hierarchy structure. Inspired by Thareja et al., who modeled conceptual progression using the hierarchical structure of academic textbooks to detect PRs [17]. We applied a similar strategy in our context

by leveraging the course hierarchy in CourseMapper, to define the CourseMapper Hierarchy (**CMH**) criterion, which captures PRs among concepts within the course hierarchy. In CourseMapper, courses are structured hierarchically, encompassing multiple topics, each of which consists of several channels that, in turn, contain various LMs. This hierarchical organization can be leveraged to infer PRs, as the arrangement of topics, channels, and LMs follows a sequential order. Specifically, channels that need to be accessed earlier are positioned before those intended for later stages of learning. By exploiting this structured progression, we can systematically derive PRs between different concepts within one course.

$$ConTop(c_1, c_2) = \begin{cases} 1, & \text{if } Top(c_1) > Top(c_2) \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

In this context, $Top(c_1)$ represents the topics or channels within one course in which c_1 appears within a learning material in CourseMapper. $ConTop(c_1, c_2)$ is assigned a value of 1 if the channel in which c_1 is mentioned precedes the channel that mentions c_2 . In this case, the likelihood of c_1 being a prerequisite for c_2 increases significantly.

Text-based Features: This approach utilizes BERTopic combined with Shannon’s entropy (**BERTropy**) to infer prerequisite connections [5]. In this context, a topic is defined as a theme discussed in the Wikipedia abstract of a given concept. For example, the abstract of the *Machine Learning* article may include topics such as *Artificial Intelligence* and *Statistical Methods*. The underlying intuition of this approach is that more advanced concepts tend to be more specialized and therefore cover fewer topics, whereas general concepts span a broader set of topics. To operationalize this, BERTopic is employed to extract topics from the Wikipedia abstract of a concept c . The model assigns a probability to each topic, indicating the degree to which the topic is present in the abstract. Let $P(x)$ denote the probability assigned by BERTopic that topic x appears in the abstract of concept c :

$$P(x) = \text{BERTopic}(x, c) \quad (14)$$

Given these topic probabilities, Shannon entropy is used to measure the distribution of topics within the abstract:

$$EntPro(c) = - \sum_x^X P(x) \log \frac{1}{P(x)} \quad (15)$$

where X denotes the set of all identified topics and $P(x)$ represents the probability of topic x for concept c . A higher entropy score indicates that the concept is more general, while a lower entropy score suggests a more advanced and focused concept. This is applied to infer PRs between all concept pairs using entropy rule. A PR is assumed to exist only if the difference exceeds this threshold θ . In

our case, the threshold is empirically selected and set $\theta = 1.33$, meaning that between 0 and 1.33, we assume no PR exists.

$$BERTropy(c_1, c_2) = \begin{cases} 1, & \text{if } EntPro(c_1) - EntPro(c_2) > \theta \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

3.2 Voting Algorithm

After applying the criteria mentioned previously to each concept pairs, a **voting algorithm** will take place determine both the presence and direction of the PR between concept pairs. First, we treat all criteria equally by assigning equal weight (1) to all of them. This avoids biasing inference toward any single feature and ensures no criterion dominates the final decision. After that, we constructed two binary arrays (one hot encoded) (A_1, A_2) corresponds to $(c_1 > c_2)$ and $(c_2 < c_1)$, each of size 10, to examine how well each concept pair satisfies the set of predefined criteria. Each array element corresponds to a specific criterion and is assigned a value of 1 if the criterion is met or 0 if not. This means that only the criteria that are satisfied contribute to the final computation used to determine the PR, while the criteria that are not met have no influence on the result. We then separately sum the positive values in each array to quantify the total contribution of satisfied criteria for each concept pair. Following that, we compute the difference between the two sums by subtracting the total score of A_1 from that of A_2 , as detailed in the following equation.

$$S_{init} = \sum A_1 - \sum A_2 \quad (17)$$

After that, normalization takes place, which is a common practice in many applications and simplifies setting thresholds for PR inference. To this end, the final score (S_{init}) is rescaled from the original range of $[-10, 10]$ to the standardized range of $[-1, 1]$. This facilitates meaningful aggregation and supporting the inference of PRs.

$$S_{new} = 2 \cdot \left(\frac{S_{init} - S_{min}}{S_{max} - S_{min}} \right) - 1 \quad (18)$$

where S_{new} is the normalized score difference between the two arrays, A_1 and A_2 . S_{min} , in this case, would be the minimum, which is -10 and S_{max} is the maximum, 10 . S_{init} is the current score before normalization a concept pair has. After normalizing the results, we classify the results into three distinct ranges. To do this, we define a threshold on both sides, these are θ and $-\theta$. The first range is if S_{new} or the difference between the arrays is between -1 and $-\theta$ which concludes that c_2 is a prerequisite concept to c_1 . When the result is between $-\theta$ and θ , that means there is no PR between the two concepts. The last range is when S_{new} is between θ and 1 then we conclude that c_1 is a prerequisite of c_2 . The value of $\theta = 0.28$ is determined empirically through a series of experiments aimed at identifying the optimal parameter setting for our specific application as mentioned in Section 4. To illustrate this more, an example in Table 2 is

given, after examining all criteria for the concept pair *Machine Learning* (c_1) and *Supervised Learning* (c_2), we use two one-hot encoded arrays (A_1, A_2) to represent how well each concept pair satisfies a set of predefined criteria, in both directions. After that, we calculate the overall sums of the two arrays ($A_1 = 5.00$

Table 2: Voting algorithm example

Criteria	TemO	HL-A	HL-Ab	HL-RCA	RefD	IOLR	CatCon	SuperCatCon	CMH	BERTropy
A_1 ($c_1 \rightarrow c_2$)	1	1	0	0	0	0	1	1	0	1
A_2 ($c_2 \rightarrow c_1$)	0	0	1	0	1	0	0	0	0	0

and $A_2 = 2.00$), then compute the difference between them and normalize it, as illustrated in Table 3. The final score of $0.3 > 0.28$ falls between θ and 1, indicating that c_1 is prerequisite to c_2 .

Table 3: Example of voting algorithm results

Mechanism	Total A_1 ($c_1 \rightarrow c_2$)	Total A_2 ($c_2 \rightarrow c_1$)	$A_1 - A_2$	Normalized difference
Voting algorithm	5.00	2.00	3.00	0.3

4 Experiments and Results

To demonstrate the effectiveness and advantages of our proposed method, we conducted thorough experiments and a comparative analysis using two public datasets, namely the AL-CPL [18, 11, 10] and Biology datasets, which is a specialized dataset derived from textbook chapters in the Biology curriculum [17].

Threshold Optimization: The threshold is essentially a hyperparameter to control the sensitivity of the model in determining the PR. The thresholds of criteria RefD and BERTopic, as well as the proposed voting algorithm were determined empirically using the same methodology, based on the AL-CPL dataset. To this end, a range of threshold values have been used to run each approach separately and compared the results with the ground truth dataset to compute precision, accuracy, recall, and F1 score. The **RefD** method identifies PRs by analyzing how concepts reference each other within the Wikipedia link structure. Wang et al. [18] originally proposed a threshold value of $\theta = 0.04$. We replicated this in the context of our platform, CourseMapper to generate the PRs. We found out that precision peaks at $\theta = 0.6$ with a value close to 0.9. This aligned with our aim of reliably detecting PRs while minimizing false positives. The higher threshold compared to previous work is attributed to the more domain-specific concept space used in our EduKG. While prior studies operated over the full Wikipedia concept set, our system limits the scope to concepts and related

concepts extracted from AL-CPL, leading to a more focused and specialized semantic space. Similarly, **BERTropy** infers PRs between concepts by analyzing the distribution of topics covered within their respective abstracts. We calculated the distribution of topics covered across all concepts. Then, BERTropy is applied to the entire set of concepts to determine PRs among them. We empirically found that precision peaks at $\theta = 1.4$, reaching approximately 0.9. This suggests that higher thresholds are more conservative, identifying only the strongest relationships, whereas lower thresholds connect more topics but may introduce noise. For the **Voting Algorithm**, we experimentally observed that precision reaches its maximum value of 1.0 at $\theta = 0.28$, indicating complete correctness among positive predictions without false positives at this threshold. In this study, we prioritize precision because minimizing false positives is more critical than capturing all possible PRs. Incorrect prerequisite suggestions could negatively affect learning paths and curriculum design.

Experiment 1: Prerequisite Learning on the AL-CPL Dataset: To evaluate the effectiveness of our proposed model, we conducted experiments using three domains from the AL-CPL dataset: Data Mining, Physics, and Macroeconomics [18,10,11]. It labels concepts with relationships: 1 for "prerequisite," -1 for "reverse prerequisite," and 0 for no relation. This dataset validates all criteria except CMH, which requires a content hierarchy that is lost when materials are uploaded into a single channel, treating all concepts at the same level. Therefore, hierarchical information was only applied in the Biology dataset, where such structure was available. The AL-CPL was uploaded in CourseMapper, and an EduKG was generated including main concepts, related concepts, and categories. This set of related concepts is utilized to infer the likelihood of a PR between concept pairs of the learning materials in CourseMapper.

Evaluation of Individual Criteria: To evaluate the contribution of each criterion, we conducted experiments across three domains—Data Mining, Physics, and Macroeconomics. Results showed that BERTropy achieved (1), HL-Ab obtained (0.84), and RefD accomplished (0.84) notably enhanced precision in Data Mining, while HL-Ab got (1) and CatCon got (1) performed well in Physics, and CatCon excelled in Macroeconomics with value (0.86). In terms of recall, IOLR with value (0.83, 0.84, 0.91) consistently outperformed all other criteria across three domains (Data Mining, Physics, and Macroeconomics), effectively capturing a broader set of prerequisite relations. Conversely, SuperCatCon underperformed in both precision and recall, indicating the need for refinement. To assess the balance between precision and recall, F1 scores were analyzed. IOLR achieved the highest F1 scores with values (0.67, 0.55, 0.59) in all domains (Data Mining, Physics, and Macroeconomics), confirming its effectiveness in maintaining this trade-off. Accuracy was also considered and offered useful trends, despite potential sensitivity to class imbalance. For example, HL-A showed high accuracy (0.65) in Data Mining, while CatCon achieved (0.70, 0.63) strong accuracy in Physics and Macroeconomics but had lower F1 scores, suggesting it may favor majority classes while poor at finding the minority class. The variation in

criterion performance across domains can be attributed to differences in concept structure and semantic clarity. For instance, Data Mining and Physics datasets often contain well-defined, hierarchical, and technically specific concepts, which align well with criteria that leverage structural or semantic features. In contrast, the Macroeconomics domain tends to be more abstract and less formally structured, making it harder for some methods to identify clear PRs. In line with prior work, we place greater emphasis on precision, as the reliability of inferred PRs is critical for downstream tasks and system utility.

Comparative Results and Analysis: We compared our proposed model against six baseline methods—RefD [9], Active Learning [11], AdaBoost [22], WikiCPRL [21], Neural Network [14], and a Linguistically-Driven Strategy [13]—to evaluate its effectiveness in capturing PRs without relying on labeled data. As shown in Table 4, we assessed our voting algorithm across three domains: Data Mining, Physics, and Macroeconomics. Compared to existing baselines, our proposed Voting Algorithm consistently achieves higher precision but lower recall and F1 scores, reflecting its focus on reliably extracting only the strongest edges. For example, in Data Mining, our method reaches perfect precision (1.0) while RefD, Active Learning, and AdaBoost achieve lower precision (0.51, 0.80, 0.85) but higher recall (0.76, 0.73, 0.85) and F1 scores (0.61, 0.67, 0.85). Similarly in Physics, Bagging-based AdaBoost achieves higher precision (0.87) than our method (0.76), while other baselines such as RefD and WikiCPRL show lower precision (0.49 and 0.43, respectively) compared to our method, but higher recall (0.49, 0.90). The same trend is observed for the Neural Network and the Linguistically-Driven Strategy, which yield higher F1 scores (0.65 and 0.81 in Data Mining, respectively, vs. 0.42 for our method) due to their stronger recall. This pattern shows that while baselines recover more links overall, our method prioritizes the precision and reliability of inferred PRs—critical for building robust learning paths—since, in educational settings, misleading a learner with an incorrect prerequisite (false positive) is more harmful than omitting some prerequisites.

However, baseline models perform better in recall, which we attribute to our precision-focused threshold selection in criteria such as RefD and BERTropy. These thresholds were tuned to prioritize reliability over quantity, resulting in fewer but more accurate inferred relationships. This can be attributed to differences between domains in concept structure and semantic clarity as mentioned earlier. Despite the lower recall, the proposed voting algorithm which integrates various feature-based methods achieves competitive performance without requiring any labeled training data. This makes it particularly well-suited for real-world settings where labeled prerequisite data is limited or difficult to obtain. Although supervised models often outperform unsupervised ones due to access to labeled data, manual labeling is costly, time-consuming, and prone to inconsistency. Most unsupervised methods evaluate criteria independently rather than integrating them for more effective PR identification. Our unsupervised multi-criteria approach addresses this gap with a scalable, flexible voting algorithm for inferring PRs, particularly in diverse educational domains lacking annotations.

Table 4: Performance Comparison Across AL-CPL Datasets

	Data Mining				Physics				Macroeconomics			
Criteria	P	R	F1	Acc.	P	R	F1	Acc.	P	R	F1	Acc.
TemO	0.47	0.31	0.37	0.40	0.36	0.42	0.39	0.47	0.46	0.71	0.56	0.53
HL-A	0.8	0.53	0.64	0.65	0.58	0.34	0.43	0.64	0.58	0.34	0.42	0.62
HL-Ab	0.84	0.22	0.35	0.53	1	0.13	0.23	0.65	0.62	0.08	0.14	0.60
HL-RCA	0.53	0.53	0.53	0.45	0.42	0.67	0.52	0.49	0.39	0.43	0.41	0.48
RefD	0.84	0.22	0.35	0.53	0.72	0.21	0.33	0.65	0.69	0.15	0.24	0.62
IOLR	0.56	0.83	0.67	0.53	0.41	0.84	0.55	0.45	0.44	0.91	0.59	0.47
CatCon	0	0	0	0.43	1	0.25	0.40	0.70	0.86	0.12	0.21	0.63
SuperCatCon	0	0	0	0.43	0	0	0	0.60	0	0	0	0.59
BERTropy	1	0.09	0.16	0.48	0.26	0.09	0.13	0.54	0	0	0	0.59
Baselines												
RefD [9]	0.51	0.76	0.61	-	0.49	0.49	0.49	-	-	-	-	-
Active Learning [10]	0.80	0.73	0.76	-	0.8	0.6	0.7	-	-	-	-	-
Bagging-based AdaBoost [22]	0.85	0.85	0.85	-	0.87	0.87	0.86	-	-	-	-	-
WikiCPRL [21]	0.59	0.90	0.68	-	0.43	0.83	0.57	-	-	-	-	-
Neural Network [14]	0.62	0.68	0.65	-	0.53	0.62	0.57	-	-	-	-	-
Linguistically-Driven Strategy [13]	0.81	0.81	0.81	-	0.77	0.71	0.74	-	-	-	-	-
<i>Proposed - Voting Algorithm</i>	1	0.27	0.42	0.58	0.76	0.13	0.22	0.64	0.65	0.16	0.25	0.61

Experiment 2: Prerequisite Learning on the Biology Dataset: To effectively evaluate the proposed model’s ability to detect PRs and assess **TemO** criterion more reliably, as well as to measure the contribution of **CMH**, we utilized a specialized dataset derived from textbook chapters in the NCERT¹ Biology curriculum [17]. For our experiments, we selected two conceptually related chapters from NCERT science textbooks: Chapter 7 (Transportation in Animals and Plants, Class 7) and Chapter 6 (Life Processes, Class 10). These were chosen for their natural hierarchical relationship, with foundational content in Class 7 supporting advanced concepts in Class 10. Both chapters were uploaded into CourseMapper under the same course and same topic, but placed in two different sequential channels to preserve the learning hierarchy defined by the platform. We generated the EduKG for each chapter individually by extracting the main concepts from each and then expanding them using DBpedia to identify related concepts. This structure enabled us to evaluate all ten predefined criteria for detecting PRs. It is worth mentioning, when a new learning material is added, the system computes the EduKG for this new material, without needing to reprocess the entire EduKG. In the absence of ground truth in the original dataset, 186 concept pairs were manually annotated by domain experts (PhDs and medical doctors) as either prerequisite or not.

Evaluation of Individual Criteria and Comparative Results: Biology dataset is used to evaluate each individual criterion using four standard metrics: precision, recall, F1-score, and accuracy (Table 5). HL-Ab and HL-A achieved the highest precision, effectively identifying the most reliable PRs while minimizing false positives. In contrast, IOLR and CMH attained the highest recall, capturing a broader set of dependencies at the expense of tolerating some false pos-

¹ <https://ncert.nic.in/textbook.php>

itives. Notably, IOLR and CMH also achieved the highest F1-scores, indicating a strong balance between precision and recall. In terms of accuracy, BERTropy performed best (0.75), followed closely by IOLR and CMH, reinforcing their robustness in distinguishing between prerequisite and non-prerequisite pairs. The success of CMH, in particular, highlights the value of structured educational content and hierarchical information in inferring PRs. In this dataset, the placement of materials across channels is well-defined, enabling CMH to leverage this ordering effectively. By contrast, CatCon and SuperCatCon depend on the category hierarchy in DBpedia, and TemO relies on the sequential appearance of concepts in text, both of which are less consistent in the current dataset. Their effectiveness may improve by incorporating additional chapters. In comparison, our proposed voting algorithm achieved competitive results, although it did not surpass the best-performing individual criteria such as CMH and IOLR. With moderate precision and recall, it provided a balanced yet limited performance compared to these criteria when applied independently. The effectiveness of the voting algorithm is largely supported by strong individual contributors like CMH and IOLR. CMH achieved high F1-scores due to its reliance on hierarchical signals, while IOLR benefited from the ratio of in-outbound links—both features being particularly well captured in the Biology dataset. These findings highlight the importance of integrating hierarchical information. The defined criteria are deduced from Wikipedia articles, DBpedia hierarchy, textual information, content order, and platform hierarchy to infer PRs, making it applicable across diverse educational platforms.

Table 5: Performance Comparison Across Biology dataset

Criteria	Biology			
	P	R	F1	Acc.
TemO	0.0	0.0	0.0	0.42
HL-A	0.91	0.32	0.48	0.58
HL-Ab	1.0	0.06	0.12	0.45
HL-RCA	0.87	0.23	0.35	0.51
RefD	0.75	0.10	0.17	0.45
IOLR	0.58	1.0	0.74	0.58
CMH	0.58	1.0	0.74	0.58
CatCon	0.0	0.0	0.0	0.42
SuperCatCon	0.0	0.0	0.0	0.42
BERTropy	0.55	0.65	0.70	0.75
<i>Proposed - Voting Algorithm</i>	0.48	0.35	0.41	0.40

5 Conclusion and Future Work

Educational Knowledge Graphs (EduKGs) capture diverse entities and connections in educational systems, providing a structured view of key concepts

from learning materials. Prerequisite relationships (PRs) are essential for defining learning sequences but are often missing in EduKGs. However, automatic methods for PR identification depend on domain-specific labeled data and static features, limiting scalability. Existing feature-based methods often ignore how semantic and structural information interact, rely often on concept order, and assess criteria individually. To address these gaps, we proposed an unsupervised method that combines ten criteria using a voting algorithm to infer PRs without using labeled data. Experiments on two benchmark datasets show that our proposed approach achieves higher precision than baselines and reliably extracts PRs across domains. In future work, we aim to extend our PR identification framework with more criteria. We will also experiment with hybrid and LLM-based approaches for more effective PR identification. Additionally, adaptive thresholding methods, learned per domain and combined with a weighting approach, could be adopted to further enhance the robustness and accuracy of PR inference.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ain, Q.U., Chatti, M.A., Bakar, K.G.C., Joarder, S., Alatrash, R.: Automatic construction of educational knowledge graphs: a word embedding-based approach. *Information* **14**(10), 526 (2023)
2. Ain, Q.U., Chatti, M.A., Joarder, S., Nassif, I., Wobiwo Teda, B.S., Guesmi, M., Alatrash, R.: Learning channels to support interaction and collaboration in coursemapper. In: *Proceedings of the 14th International Conference on Education Technology and Computers*. pp. 252–260 (2022)
3. Alatrash, R., Chatti, M.A., Ain, Q.U., Fang, Y., Joarder, S., Siepmann, C.: Conceptgcn: Knowledge concept recommendation in moocs based on knowledge graph convolutional networks and sbert. *Computers and Education: Artificial Intelligence* **6**, 100193 (2024). <https://doi.org/https://doi.org/10.1016/j.caeai.2023.100193>, <https://www.sciencedirect.com/science/article/pii/S2666920X23000723>
4. Bai, Y., Liu, Z., Guo, T., Hou, M., Xiao, K.: Prerequisite relation learning: A survey and outlook. *ACM Computing Surveys* (2025)
5. Bai, Y., Zhang, Y., Xiao, K., Lou, Y., Sun, K.: A bert-based approach for extracting prerequisite relations among wikipedia concepts. *Mathematical Problems in Engineering* **2021**(1), 3510402 (2021)
6. Gasparetti, F., De Medio, C., Limongelli, C., Sciarrone, F., Temperini, M.: Prerequisites between learning objects: Automatic extraction based on a machine learning approach. *Telematics and Informatics* **35**(3), 595–610 (2018)
7. Hu, H., Pan, L., Ran, Y., Kan, M.Y.: Modeling and leveraging prerequisite context in recommendation. *arXiv preprint arXiv:2209.11471* (2022)
8. Hu, X., He, Y., Sun, G.: Active learning for concept prerequisite learning in wikipedia. In: *Proceedings of the 2021 13th International Conference on Machine Learning and Computing*. pp. 582–587 (2021)

9. Liang, C., Wu, Z., Huang, W., Giles, C.L.: Measuring prerequisite relations among concepts. In: Proceedings of the 2015 conference on empirical methods in natural language processing. pp. 1668–1674 (2015)
10. Liang, C., Ye, J., Wang, S., Pursel, B., Giles, C.L.: Investigating active learning for concept prerequisite learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018)
11. Liang, C., Ye, J., Zhao, H., Pursel, B., Giles, C.L.: Active learning of strict partial orders: A case study on concept prerequisite relations. arXiv preprint arXiv:1801.06481 (2018)
12. Manrique, R., Pereira, B., Mariño, O.: Exploring knowledge graphs for the identification of concept prerequisites. *Smart Learning Environments* **6**(1), 21 (2019)
13. Miaschi, A., Alzetta, C., Cardillo, F.A., Dell’Orletta, F.: Linguistically-driven strategy for concept prerequisites learning on italian. In: Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications. pp. 285–295 (2019)
14. Moggio, A., Parizzi, A.: Unige se@ prelearn: Utility for automatic prerequisite learning from italian wikipedia. *EVALITA Evaluation of NLP and Speech Tools for Italian-December 17th, 2020* p. 376 (2020)
15. Pan, L., Li, C., Li, J., Tang, J.: Prerequisite relation learning for concepts in moocs. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1447–1456 (2017)
16. Sayyadiharikandeh, M., Gordon, J., Ambite, J.L., Lerman, K.: Finding prerequisite relations using the wikipedia clickstream. In: Companion Proceedings of The 2019 World Wide Web Conference. pp. 1240–1247 (2019)
17. Thareja, R., Garg, R., Baghel, S., Dwivedi, D., Mohania, M., Kulshrestha, R.: Auto-req: Automatic detection of pre-requisite dependencies between academic videos. In: Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023). pp. 539–549 (2023)
18. Wang, S., Ororbia, A., Wu, Z., Williams, K., Liang, C., Pursel, B., Giles, C.L.: Using prerequisites to extract concept maps from textbooks. In: Proceedings of the 25th acm international on conference on information and knowledge management. pp. 317–326 (2016)
19. Xia, J., Li, M., Tang, Y., Yang, S.: Course map learning with graph convolutional network based on aucm. *World Wide Web* **26**(5), 3483–3502 (2023)
20. Xiao, K., Fu, Y., Zhang, J., Tianji, W.: A hybrid approach for discovering concept prerequisite relations in wikipedia. In: 2022 9th International Conference on Behavioural and Social Computing (BESC). pp. 1–5. IEEE (2022)
21. Xiao, K., Li, K., Zhang, Y., Chen, X., Lou, Y.: Wikicprl: A weakly supervised approach for wikipedia concept prerequisite relation learning. In: Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data. pp. 177–192. Springer (2023)
22. Zhou, Y., Xiao, K., Zhang, Y.: An ensemble learning approach for extracting concept prerequisite relations from wikipedia. In: 2020 16th International Conference on Mobility, Sensing and Networking (MSN). pp. 642–647. IEEE (2020)
23. Zhu, Y., Zamani, H.: Predicting prerequisite relations for unseen concepts. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022) (2022)