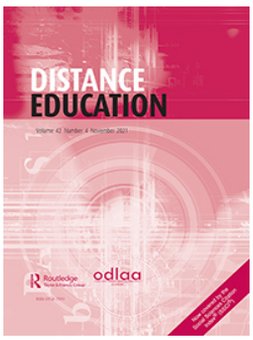


Distance Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cdie20

Toward building a fair peer recommender to support help-seeking in online learning

Chenglu Li, Wanli Xing & Walter L. Leite

To cite this article: Chenglu Li, Wanli Xing & Walter L. Leite (2022): Toward building a fair peer recommender to support help-seeking in online learning, Distance Education, DOI: 10.1080/01587919.2021.2020619

To link to this article: https://doi.org/10.1080/01587919.2021.2020619

Published online: 13 Feb 2022.

DISTANCE EDUCATION
https://doi.org/10.1080/01587919.2021.2020619

Routledge
Taylor & Francis Group

Toward building a fair peer recommender to support help-seeking in online learning

Chenglu Li[a], Wanli Xing[a], and Walter L. Leite[b]

[a]School of Teaching & Learning, College of Education, University of Florida, Gainesville, USA; [b]School of Human Development and Organizational Studies in Education, College of Education, University of Florida, USA

ABSTRACT

Help-seeking is a valuable practice in online discussion forums. However, the asynchronicity and information overload of online discussion forums have made it challenging for help seekers and providers to connect effectively. This study formulated a new method to provide fair and accurate insights toward building a peer recommender to support help-seeking in online learning. Specifically, we developed the fair network embedding (Fair-NE) model and compared it with existing popular models. We trained and evaluated the models with a large dataset consisting of 187,450 discussion post-reply pairs by 10,182 Algebra I online learners from 2015 to 2020. Finally, we examined models with representation fairness, predictive accuracy, and predictive fairness. The results showed that the Fair-NE can achieve superior fairness in genders and races while retaining competitive predictive accuracy. This study marks a paradigm change from previous investigation and evaluation of fair artificial intelligence to proactively build fair artificial intelligence in education.

ARTICLE HISTORY
Received 23 July 2021
Accepted 15 December 2021

KEYWORDS
fair artificial intelligence; peer recommenders; help-seeking; online discussion forums

Introduction

Help-seeking in online discussion forums has been demonstrated to be an important learning strategy (Chao et al., 2018). It is a strategy in self-regulated learning and goes beyond stating the need for assistance. Studies have shown that help-seeking can positively correlate with students' self-efficacy (Chyr et al., 2017), motivation (Melrose, 2006), and learning outcomes (Parnes et al., 2020). Among different sources of support for help-seeking, such as instructional teams and intelligent agents, peer support stands out given its collaborative nature, which enhances learning through knowledge exchange (Kear, 2004). However, seeking help among peers in online discussion forums can be challenging. On one hand, students can find it challenging to know whom to reach out to in an unfamiliar asynchronous environment (Labarthe et al., 2016). On the other hand, help providers can be overwhelmed with disorganized discussion forums swamped with threads, for example, massive open online course

CONTACT Wanli Xing wanli.xing@coe.ufl.edu

(MOOC) forums (Shaw, 2019), not knowing where to help. Consequently, the lack of effective connections between help seekers and providers in online discussion forums can impede students' help-seeking practices, yielding a potential threat to students' sense of belonging and engagement (Allen & Kern, 2017).

Recent studies have shown an increasing interest in building peer recommenders with artificial intelligence (AI) to automatically connect help seekers with providers, given the potential to initiate sustainable and reciprocal conversations among peers. A peer recommender in education is a system to provide students personalized social outreach to help expand their social networks and potentially assist with their learning (Garcia-Martinez & Hamou-Lhadj, 2013). There is considerable theoretical research behind such systems in education. From the perspective of help-seeking in self-regulated learning, peer recommenders assist with identifying help providers, which is an important stage in help-seeking (Cross et al., 2017). Furthermore, peer recommenders can greatly improve the chances that students will receive peer feedback and suggestions on their statements or questions by providing intelligence to suggest peers who are likely to connect. The potentially enhanced bond among students might improve students' retention and engagement with the online learning community (Labarthe et al., 2016). Moreover, the selection mechanism of a peer recommender tends to prioritize peers who share similar interests in discussion topics, strengthening the ties between students and leading to sustainable and reciprocal discourses (Potts et al., 2018).

From the algorithm perspective, the core component of a peer recommender is the prediction of students' possible connections in a network, named *link prediction*. There are two popular approaches to building link prediction models for AI-based peer recommenders. One utilizes structural similarity with traditional social network analysis (SNA) to suggest peers who would establish connections with help seekers (e.g., Hansen et al., 2019; Sunar et al., 2017; Yang et al., 2018). A new approach called *network embedding* has been shown as a strong candidate to build peer recommenders (Chai et al., 2019; Xiao et al., 2021). Network embedding utilizes deep neural networks to represent nodes in a graph with latent vectors such that neighboring nodes would have high similarity scores (Nelson et al., 2019). Researchers have found that network embedding has advantages of computational efficiency and predictive accuracy over prior link prediction algorithms (Grover & Leskovec, 2016; Nelson et al., 2019; Xu et al., 2020).

Current studies on building AI-based peer recommenders have focused on improving prediction performance. Namely, researchers have examined enhanced or innovative models to make recommendations more accurate and have witnessed some successes. However, little is known as to whether these AI systems will treat students fairly. There have been reports that AI models can be biased against demographic factors such as gender and race across domains: education (e.g., Riazy & Simbeck, 2019), business hiring (e.g., van den Broek et al., 2020), and medicine (e.g., Esteva et al. 2017). In the case of peer recommenders, students might form communities of specific demographics that could lead to AI bias. For example, Caucasian students dominantly interact with other Caucasian students because they come from the same school where minority students are scarce. Trained with such a dataset, the recommender system can reinforce the status quo and not give students opportunities to establish

diversified connections that can be equally helpful. Some conceptual studies (Baker & Hawn, 2021; Kizilcec & Lee, 2020; Marcinkowski et al., 2020) and assessment studies (Riazy & Simbeck, 2019; Sha et al., 2021; Yu et al., 2020) have discussed and evaluated whether educational AI systems can deliver fair intelligence for learners. However, there has not been a focused investigation into how to increase AI fairness in educational peer recommenders strategically.

In this study, we aimed to fill this gap by taking the first step by developing a new algorithm that can improve the fairness of AI-based peer recommenders for education, inspired by the work of Buyl and De Bie (2020). Furthermore, we intended to examine and synthesize strategies to evaluate the AI fairness of peer recommenders. Specifically, our study is rooted within the discussion forum in Algebra Nation (https://www.algebranation.com/), an online math learning platform that originated in Florida. Using Algebra Nation as the research context, we created a network embedding model that can fairly represent students' social attributes as latent vectors while being competitive on link prediction accuracy. We evaluated the fair network embedding (Fair-NE) model with existing widely adopted network embedding models—Node2Vec (Grover & Leskovec, 2016) and FairWalk (Rahman et al., 2019). Details of these models are discussed in the Methods section. The network embedding models were built to power link prediction models in the discussion forum to recommend students who are likely to connect with a particular help seeker. The results suggest that Fair-NE can learn a fair representation of students, positively affecting link prediction's fairness and accuracy. This study marks a paradigm change from previous investigation and evaluation about fair AI to proactively construct fair AI in education.

Background

Previous studies on AI-based peer recommenders in education

AI-based peer recommenders can be constructed twofold: (1) a link prediction approach to suggest peers who are likely to establish connections with a target student through graph theory, and (2) a student-modeling approach to recommend peers with similar behaviors or backgrounds. The former provides a list of candidates for connections, and the latter scores the candidates to allow further granular matching. For link prediction, a majority of studies have used social network topologies such as node attributes (e.g., Hansen et al., 2019), edge attributes (e.g., Sunar et al., 2017), and community detection (e.g., Yang et al., 2018) to construct models. For example, Hansen et al. included students' closeness and betweenness centralities such as shortest connection length and involvement in multiple subcommunities in the social network to improve link prediction performance by 20%. Sunar et al. showed that the use of interaction strength in social networks could enhance link prediction performance in MOOCs. They calculated the interaction strength based on the weighted frequency of interactions among students in the discussion forum. Yang et al. (2018) extended Sunar et al. 's work and adopted time-series models and social network's topological features such as common neighborhoods to have significantly improved link prediction's performance in a MOOC discussion forum.

In terms of student modeling, researchers have examined a variety of attributes such as learner profiles (e.g., Garg & Goel, 2021; Sun et al., 2020), learning behaviors

(Xu & Yang, 2015), and learning progress (e.g., Bouchet et al., 2017) to recommend peers highly similar in those aspects. For example, Garg and Goel used students' demographic factors such as location, gender, and age to find similar peers on MOOCs. Sun et al. built their peer recommender with the Big Five personality framework (Costa & McCrae, 2008) to recommend peers with similar personality traits in an online learning platform. Xu and Yang proposed to match students who were similar in terms of discussion forum behaviors using log and text data. Bouchet et al. built a recommender that matched MOOC students with similar learning progress (e.g., modules learned and videos viewed) and compared it with a demographic-based recommender. The results suggested that students were more likely to accept recommendations and increase interactions with the demographic-based recommender.

Network embedding for link prediction in education

The studies above on link prediction (Hansen et al., 2019; Sunar et al., 2017; Yang et al., 2018) focused on the use of traditional SNA. However, recent works on link prediction have suggested that using network embedding can be advantageous on both computational efficiency and predictive accuracy over traditional SNA (Nelson et al., 2019; Xu et al., 2020). Network embedding uses deep neural networks to represent students' social attributes in a graph with numeric vectors. Unlike traditional SNA, which represents students in a single dimension, network embedding encodes students' characteristics in a network with multidimensional representations, which can be used to conduct further analysis in different angles: link prediction (e.g., peer recommenders, Xiao et al., 2021), clustering (e.g., learning community detection, Wu et al., 2020), and classification (e.g., domain knowledge acquisition, Abu-Salih et al., 2021).

Although the advantages of network embedding, such as superior predictive accuracy and versatile analysis, have been demonstrated in previous studies (Nelson et al., 2019; Xu et al., 2020), only a few studies have examined the affordances of network embedding in educational contexts (Chai et al., 2019; Xiao et al., 2021). Chai et al. utilized network embedding to recommend potential empathizers in an online healthcare community. Their results showed that the network embedding model could achieve the best predictive accuracy among other benchmark models with SNA and user modeling. Xiao et al. created a network embedding method that handled different types of relationships for peer recommenders on MOOCs, and their results showed great performance improvements over baseline models. Disregarding different methods to build peer recommenders, most studies in education have focused mainly on improving recommenders' predictive accuracy. In contrast, little has been done to evaluate fairness and proactively reduce bias in educational AI systems. The following section discusses the current landscape of fair AI in education.

Research on fair AI in education

Studies of fair AI have focused on discussing, evaluating, and reducing bias related to AI algorithms. AI algorithms might be neutrally designed, but input data passed to algorithms can be biased, thus causing algorithmic biases (Zimmer et al., 2019).

Algorithmic biases are unwanted behaviors of algorithms that embed the hidden values of humans (Beer, 2017). Algorithmic biases in AI have been widely identified in various fields. For example, van den Broek et al. (2020) found that hiring algorithms could favor candidates with an outgoing personality, while in reality, the top performers in their sample were likely those with an introverted personality. Esteva et al. (2017) found skin cancer diagnosis algorithms could be biased against dark-skinned people, underestimating the chances of dark-skinned people developing skin cancer. AI in education will not be warranted from algorithmic biases.

A recent educational research initiative on fair AI has concentrated on conceptualizing and assessing algorithmic bias, aiming to raise awareness of AI fairness in educational settings. For example, Kizilcec and Lee (2020) undertook an in-depth examination of mathematical concepts on fairness that could be applied to educational research. They discussed the benefits and harms of the probabilities of AI models being able to correctly and incorrectly identify positive cases in educational contexts. Building on Kizilcec and Lee's study, Baker and Hawn (2021) discussed the source of AI bias in education. They investigated that some types of bias could be handled before and during data collection. Measurement bias, for example, can be mitigated by ensuring construct validity. They also examined bias that could require extra attention in AI model training: historical bias (e.g., using historically biased data in combination with new data), aggregation bias (e.g., combining data from different populations), and evaluation bias (e.g., evaluating models with unrepresentative subsamples). Instead of employing the ideas of fairness from a mathematical standpoint, Marcinkowski et al. (2020) conducted a study to understand students' perceptions of AI fairness toward an AI-based college admission system. Their results revealed that individual's perceived fairness could greatly impact their opinions on applications to the college and educational AI systems.

Other studies have evaluated whether AI models could meet their fairness criteria in educational contexts. Riazy and Simbeck (2019) verified whether their AI models' prediction accuracy differed for students with disabilities. They measured the models' biases and discovered that the models had differing degrees of bias, whose prediction accuracy was, however, satisfying. Similarly, Sha et al. (2021) examined the fairness of AI-based classifiers that automatically categorized students' forum posts in a learning management system. They evaluated AI fairness against gender and language based on the absolute between-ROC area (ABROCA) developed by Gardner et al. (2019). ABROCA is the area between two receiver operating characteristic (ROC) curves from subgroups (e.g., female and male). The results showed that most AI models favored students whose first language was English. Hutt et al. (2019) also utilized ABROCA to evaluate AI models' fairness in predicting whether students would graduate on time, from the perspectives of race and socioeconomic status. Their results suggested that little bias could be detected in their AI models.

Research aim

Fairness challenges originating from AI models in educational contexts should receive more attention as data-driven systems are increasingly adopted in K-12 and higher

education to support teaching and learning (Kizilcec & Lee, 2020). Research of fair AI in education has conducted extensive conceptual discussion and empirical evaluation of AI fairness. However, little is known about what strategies researchers can adopt proactively to reduce AI bias in education. This study aimed to create and evaluate a fairness-aware algorithmic strategy to serve as the first step to inform studies on building fair AI-based peer recommenders for education. Specifically, we asked the research question: To what extent can the Fair-NE model provide intelligence to support help-seeking in a fair and accurate manner? Our hypothesis was that bias embedded in data would lead to bias in fairness-unaware models trained with such data, while fairness-aware models that specifically address data bias would achieve better fairness. To explore the research question, we built three network embedding models using existing solutions as well as ours and evaluated their representation fairness, predictive accuracy, and predictive fairness.

Methods

Research context and data

Data collection in this study received Institutional Review Board approval from the University of Florida. We collected data from the online math learning platform Algebra Nation. A total of 500,000 students across six states use the platform every year (University of Florida, n.d.). We randomly sampled 10,500 Algebra I learners from the MySQL database of Algebra Nation along with all the posts generated by them from 2015 to 2020. After removing students without demographic information and their posts, the final dataset consisted of 187,450 discussion post-reply pairs by 10,182 students. Although Algebra Nation also offers other math courses such as Geometry and Algebra II (University of Florida, n.d.), this study targeted Algebra I students because this was the first course offered on Algebra Nation. A majority of students enrolled in Algebra Nation took the Algebra I course; thus, making the study potentially beneficial to many students. Of the participants, 48.21% ($n = 4{,}909$) were females and 51.79% ($n = 5{,}273$) were males. In terms of race, 66.86% ($n = 6{,}808$) of Caucasian, 15.57% ($n = 1{,}586$) of Black or African-American, 7.13% ($n = 726$) of Asian, 1.10% ($n = 112$) of American Indian, 0.17% ($n = 17$) of Hawaiian or other Pacific Islander, 4.70% ($n = 479$) of two or more races, and 0.11% ($n = 11$) of unspecified races.

Fairness comparison groups

There have been various reports on the underrepresentation of female and minority students in the context of science, technology, engineering, and math (STEM) education (Office of Civil Rights, 2016; Smith et al., 2013). Therefore, we chose to examine the fairness of the peer recommender regarding gender and race. We further categorized the race attribute into a binary value following the National Science Foundation (2017) report on STEM education: overrepresented and underrepresented. Specifically, we coded Caucasian and Asian students as overrepresented ($n = 7{,}534$). On the contrary, we coded students who were Black or African-American, American Indian,

Hawaiian or other Pacific Islander, multiple racial identities, and unspecified races as underrepresented ($n = 2{,}648$).

Fair network embedding and benchmarks

For the network embeddings, we benchmarked Fair-NE with Node2Vec (Grover & Leskovec, 2016) and FairWalk (Rahman et al., 2019). Node2Vec was inspired by the widely applied natural language processing algorithm Word2Vec (Mikolov et al., 2013). Figure 1 demonstrates the learning process of Node2Vec to represent nodes with numeric information compared to Word2Vec. Word2Vec represents words with numeric vectors such that similar words (e.g., "love" and "enjoy") or words that often exist in the same context (e.g., "algebra" and "nation") can be computed with high similarities. The mechanism of Word2Vec helps machines understand and retain the contextual meanings of data, which can also be adopted in social networks for encoding nodes. Similarly, Node2Vec represents nodes (e.g., students) in a social network with latent vectors to effectively capture information of similar nodes or nodes that are (un)directly connected, thus understanding the relationship between nodes. In Node2Vec, network nodes are analogous to words in Word2Vec. Node2Vec samples sentences or sequences of nodes by iteratively selecting a starting node randomly, visiting other connected nodes following the edges, and outputting the visited nodes as a sequence when a predefined sequence length has been met. This sampling process of node sequence is called *random walk*. Latent vectors of nodes are then extracted from a neural network's hidden layer trained with the constructed sequences of nodes. Although Node2Vec is widely accepted in the network embedding community, it is fairness-unaware. To help Node2Vec fairly encode students' network information, Rahman et al. (2019) proposed using FairWalk. FairWalk is almost identical to Node2Vec except that its sampling strategy is modified to ensure that each node with a specific sensitive attribute (e.g., race) has an equal chance of being selected.

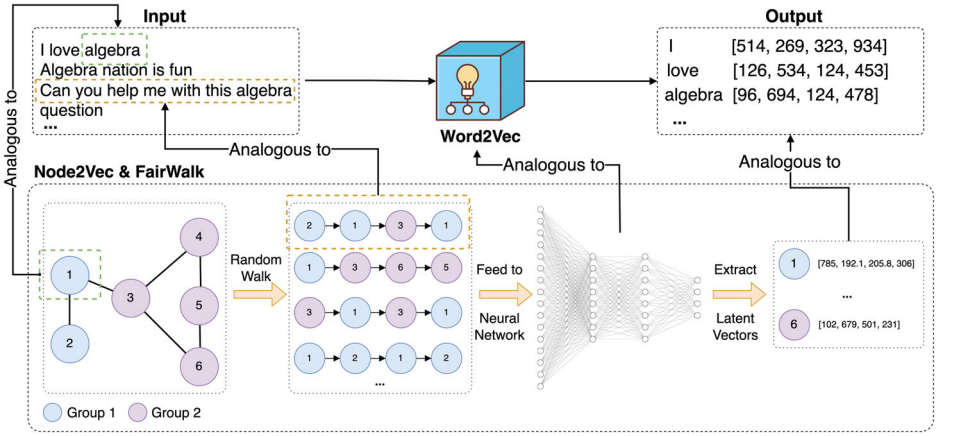


Figure 1. Illustration of the learning mechanism of Node2Vec and FairWalk compared to Word2Vec.

Table 1. Explanations of the notations used in Equation 1.

Notations	Explanations
$P(G\|X)$:Posterior	Probability distribution of seeing network G when latent vector X is known (or set to a fixed value).
$P(X\|G)$: Likelihood	Probability distribution of latent vector X of different values when network G is known. Conceptually, distances between nodes will be different as X changes. This distribution can be interpreted as the probability distribution of pairwise distance combinations.
$P(G)$: Prior	Probability distribution of network G with different structures. This is defined by users or inferred from data to provide knowledge of how densely connected a network is and how many incoming and outgoing connections each node has.
$P(X)$: Marginal likelihood	Probability distribution of latent vector X of different values. This can be calculated with $P(X\|G)$ and $P(G)$.

Rahman et al. showed that the modified sampling of node sequences could help achieve fairer results.

Fair-NE is different from Node2Vec and FairWalk in two ways. Firstly, we used a Bayesian approach to solve for:

$$P(GX) = \frac{P(X|G)P(G)}{P(X)} \tag{1}$$

where X is each node's latent vector that can maximize the likelihood of seeing a given network G's structure (e.g., how nodes are connected). Table 1 explains the notations used in Equation 1. Conceptually, the latent vector consists of coordinates of network nodes in a high dimensional space (e.g., 50 dimensions). Although values in a single dimension might not yield informative insights on nodes' relationships, the combination of values in multiple dimensions can distill knowledge about the neighborhoods of each node (Nelson et al., 2019). The goal of Equation 1 is to iteratively adjust the coordinate values of nodes such that researchers can be confident that nodes will be connected in the way demonstrated in the training data. Latent vector X is initialized with random values. Fair-NE will learn to adjust X repetitively to minimize the distances between neighboring or connected nodes and to maximize the probability of reproducing the given network structure.

The possibility of defining a prior distribution allows the integration of individual or external information into model learning. Thus, the latent vectors will need to represent only information not already captured by the prior distribution. For example, researchers can inform Fair-NE how many neighbors and genders each student of a specific gender is connected to and how many actual and potential connections each gender has in general. Such information can be inferred from the network structure using the existing discussion interaction data. The learned latent vectors in Fair-NE will thus be greatly debiased as there is no need for the latent vectors to know students' sensitive information. Fair-NE also adopts a different sampling strategy than Node2Vec and FairWalk. In Node2Vec and FairWalk, negative samples (nodes that are not likely to be connected) are drawn simplistically. Specifically, nodes that are outside a window (a sequence length defined by the researchers) will be sampled randomly as negative. However, in Fair-NE, negative sampling is achieved with the robust negative sampling strategy developed by Armandpour et al. (2019), which can yield better predictive accuracy for link prediction.

Fairness metrics

In this study, we examined fairness from two dimensions. First, we studied the representation fairness of the latent vectors from network embeddings. We used representation bias (RB) to understand if sensitive attributes such as gender and race can be predicted through the latent vectors. An ideal result would show that latent vectors from network embeddings will provide little information on students' gender and race. We utilized balance score (BS) to determine whether clusters trained with the latent vectors can fairly assign communities to students. We expected to see more balanced communities with fairer models. Second, using equalized odds (EO) and ABROCA, we investigated the predictive fairness of link prediction models that took the latent vectors as input. For desirable predictive fairness, we expected that students' genders and races would not influence the link prediction model to under- or overestimate their probabilities to connect. For example, it would not yield more correct connections for males than females. The following paragraphs discuss the metrics for representation fairness and predictive fairness in detail.

RB

Zemel et al. (2013) defined RB as the area under the ROC curve, also known as AUC, when using latent vectors to predict sensitive attributes (e.g., gender). A ROC curve visualizes how the efficiency of prediction changes with different probability threshold values, above which predictive models will yield positive prediction, otherwise negative. AUC measures predictive accuracy with a point value to allow a direct comparison between models. RB has been widely adopted in the fairness examination of models that utilize latent vectors such as natural language processing (Caliskan et al., 2017) and network embeddings (Bose & Hamilton, 2019). To evaluate RB, we used the latent vectors of network embeddings as input to construct a classifier that predicts students' gender or race. Conceptually, latent vectors of network embedding models are fair when RB is close to 0.5 since an AUC of 0.5 suggests a random classifier; researchers cannot infer students' sensitive information from latent vectors. Such desensitized information can contribute to the fairness of spatial inference, such as community detection with clustering techniques.

BS

Chierichetti et al. (2017) developed BS to indicate the fairness of clustering results. For example, a fair cluster should have a female-to-male ratio close to 0.92 (e.g., 0.48:0.52) if the female-to-male population ratio is 0.92. BS measures how well-proportioned are binary values (e.g., female and male) of a demographic variable across multiple clusters. Having such a fair representation of sensitive attributes such as race and gender is vital to apply clustering results to support students' learning (Quy et al., 2021). We conducted clustering with students' latent vectors and used BS to evaluate whether fairness-aware models could contribute to achieving fair representation. Figure 2 demonstrates an example of calculating the balance score for clustering.

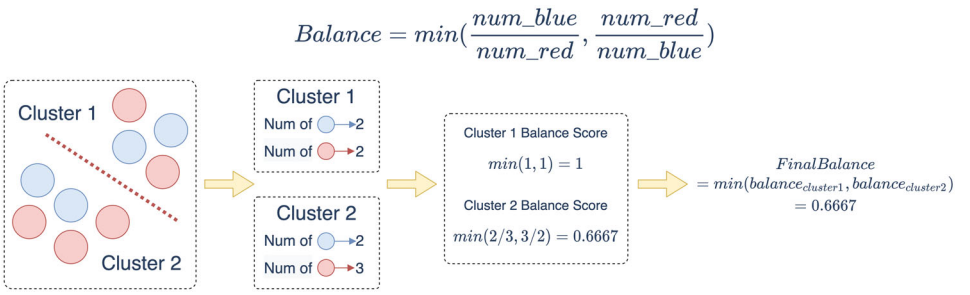


Figure 2. Balance score calculation.

EO

Hardt et al. (2016) defined EO to provide researchers with insights on predictive fairness. Researchers can use EO to address individual fairness such that a predictive model would yield similar results for students with a similar background. A variety of AI fairness evaluation studies in education have adopted EO to help researchers identify the risk of AI bias (e.g., Li et al., 2021; Riazy & Simbeck, 2019; Yu et al., 2020). Conceptually, a perfect EO suggests a model will not make more mistakes or correct judgments for a specific demographic group. EO is formulated as:

$$\mathrm{P}\left(\hat{Y} = 1 | \mathrm{A} = 0, \mathrm{Y} = 1\right) = \mathrm{P}\left(\hat{Y} = 1 | \mathrm{A} = 1, \mathrm{Y} = 1\right) \tag{2}$$

$$\mathrm{P}\left(\hat{Y} = 1 | \mathrm{A} = 0, \mathrm{Y} = 0\right) = \mathrm{P}\left(\hat{Y} = 1 | \mathrm{A} = 1, \mathrm{Y} = 0\right) \tag{3}$$

where $\hat{Y}$ is the predicted outcome of the model, Y is the ground-truth outcome from the dataset, and A is the comparison group (e.g., gender). EO demands the same true positive rate (TPR; see Equation 2) and false positive rate (FPR; see Equation 3) between groups (e.g., female vs. male). However, achieving the same TPR and FPR can be challenging in real life; therefore, we defined a scoring function of EO as:

$$\gamma(\hat{Y}) = \max\left(\left|\mathrm{FPR}_{\mathrm{group1}} - \mathrm{FPR}_{\mathrm{group2}}\right|, \left|\mathrm{TPR}_{\mathrm{group1}} - \mathrm{TPR}_{\mathrm{group2}}\right|\right) \tag{4}$$

taking the maximum between the absolute values of TPR differences and FPR differences between groups. In Equation 4, lower values indicate fairer results. A model might yield the same percentage of wrong predictions for female and male students, which might make it seem fair. However, researchers also need to check whether the model shows a similar predictive correction rate between females and males. If there were a great difference in the correction rate, the model would still be deemed biased.

ABROCA

Gardner et al. (2019) developed ABROCA to improve upon EO. Similar to EO, lower ABROCA values suggest a similar TPR and FPR between groups, thus indicating fairer results. However, EO requires a probability threshold to determine positive and negative predictions. Although the threshold of 0.5 is commonly used with AI models, there are flexibilities of choosing different values (e.g., Esposito et al., 2021). ABROCA, on the other hand, calculates the absolute area between two ROC curves, which considers the predictive performance within the entire range of thresholds from 0 to 1. ABROCA is the first AI fairness metric developed in the educational research

community and has been adopted in fairness evaluation studies in education (Hutt et al., 2019; Sha et al., 2021). However, ABROCA has not yet attracted enough attention in the general fair AI community. Therefore, this study used both EO and ABROCA to provide rich benchmarks for future studies.

Data analysis

Descriptive statistics

We first computed descriptive statistics on the discussion data to help us analyze if the dataset might be biased. The descriptive statistics consisted of the distributions of gender and race, the frequency of students' interactions in gender and race, and the communities that students have formed. We used network modularity to detect communities of students to examine if students have formed communities dominated by a specific gender or race. Modularity measured how densely were student connected within a community and between communities, which has been used in various educational studies on SNA (see the review of Phillips & Ozogul, 2020).

Experimental setup for model training and evaluation

We set up an experiment to train and evaluate network embedding and link prediction models. The experimental process is shown in Figure 3.

1. We split the post-reply pairs, namely, network edges, into training ($n = 149{,}960$) and testing edges ($n = 37{,}490$). We then generated an equal number of negative edges for testing and merged them with the testing positive edges, with a total of 74,980 edges.

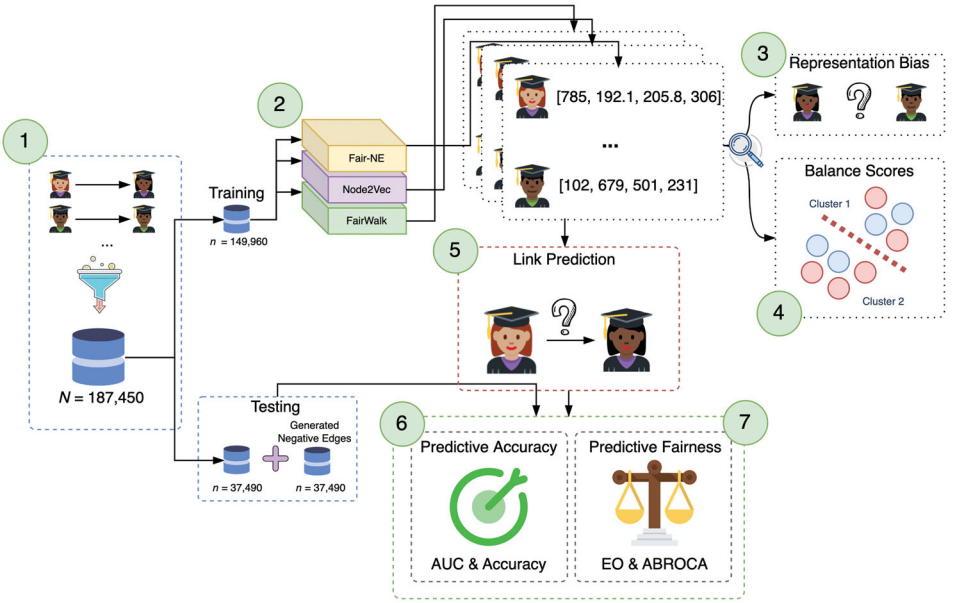


Figure 3. Experimental process.

2. We used the training edges to allow Fair-NE, Node2Vec, and FairWalk to learn latent vectors of network embeddings.
3. We then examined the representation bias of these latent vectors using a logistic regression classifier.
4. We conducted k-means clustering with the latent vectors and calculated the clustering balance scores accordingly. We tried the range from 2 to 20 as the number of clusters to examine the relationship between clusters' ability to partition students and their balance scores. We then used the popular elbow method (Kodinariya & Makwana, 2013) to determine the best number of clusters.
5. We built link prediction models with a logistic regression by using the latent vectors from the three network embedding models as input (independent variable) to predict the probability of other students connecting with a target student (dependent variable).
6. We used AUC and accuracy to determine the link prediction models' predictive accuracy. A commonly used threshold of 0.5 was adopted to determine positive and negative classes to calculate accuracy.
7. We used EO and ABROCA to determine the predictive fairness of the link prediction models. We used 2,000 linearly interpolated values in the range of 0 to 1 as the predictive thresholds to calculate EO, aiming to comprehensively understand how different thresholds of yielding positive cases would affect EO.

Results

Descriptive statistics

Figure 4 shows the distribution of students' gender and race. Figure 5 shows how frequently students with different genders and races interact. The results showed that students tended to connect with others who shared a similar demographic background. For example, there were 55,217 discussion interactions between female students whose racial identity was overrepresented, greatly outnumbering students'

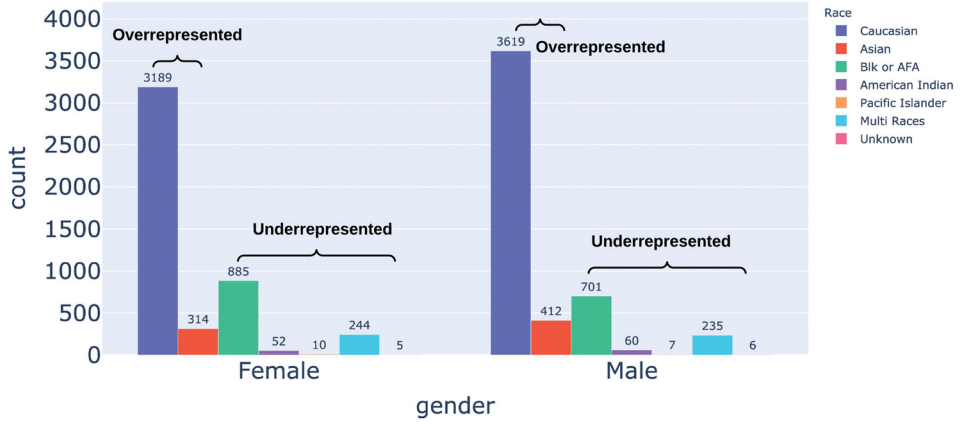


Figure 4. Distributions of gender and race in combination. *Note.* Blk or AFA = Black or African-American; Pacific Islander = Hawaiian or other Pacific Islander; Multi Races = two or more races; Unknown = unstatedspecified.

interactions with different demographic backgrounds. Figure 6 shows the communities detected by network modularity, with nodes tinted by students' gender and race. Students within each community were densely connected, while they were relatively loosely connected with students outside their community. Both Figure 6(a) and Figure 6(b) have zoomed-in example communities dominated by students of a specific demographic. It is interesting to note in Figure 6(a) that there were communities whose members were dominantly male, given the similar proportions between female (48.21%) and male (51.79%) students. A similar pattern applies to race (see Figure 6(b)), where overrepresented students heavily interacted with other overrepresented students.

Model evaluation

This section presents the results of the network embedding and link prediction models. First, we show the results of representation bias and clustering balance scores of latent vectors. Then, we present the AUC and accuracy of the link prediction models. Finally, we demonstrate the predictive fairness of the link prediction models with EO and ABROCA.

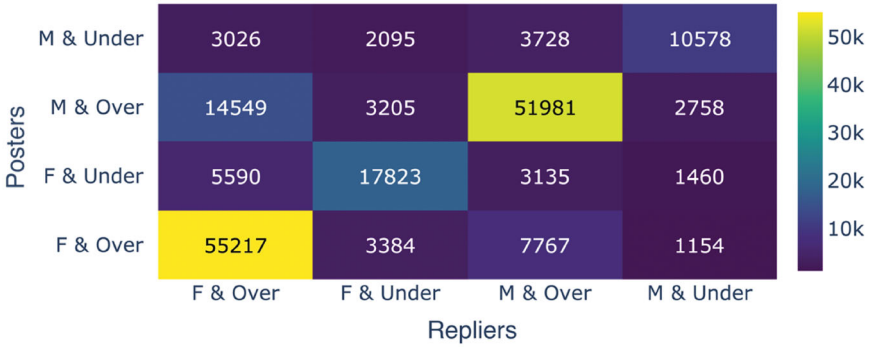
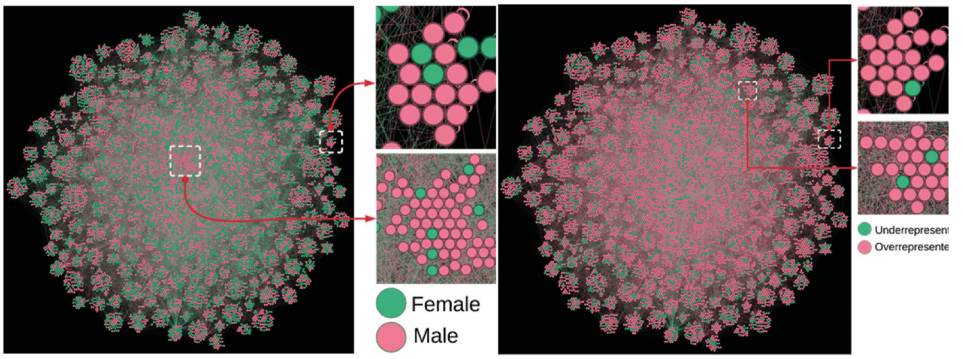


Figure 5. Frequency of students' discussion interactions through the lens of gender and race.
Note. F = females; M = males; Under = underrepresented; Over = overrepresented.



(a) Tinted by gender (b) Tinted by race

Figure 6. Communities of students' interactions on the discussion forum.

Representation bias

Figure 7 shows the representation bias of the three network embedding models in terms of gender and race. The boxplot showed that Fair-NE achieved the fairest representation among the three models. The results showed that FairWalk was the least fair when using race as the comparison group, meaning more information on students' race was encoded in the latent vectors of FairWalk.

Clustering balance scores

Figure 8 shows the clustering results of the latent vectors regarding the clusters' ability to partition students. In the analysis, the three models happened to have the same best number of clusters of 5. Figure 9(a) shows the clustering balance scores of gender, while Figure 9(b) shows those of race. In terms of gender, all three models achieved the best balance score when there were two clusters (see Figure 9(a)), while the second-best balance score was identified when there were around 5 clusters. Interestingly, Fair-NE's best balance score was located at 5 clusters for race, while the balance scores of Node2Vec and FairWalk still peaked at 2 clusters (see Figure 9(b)). Generally, Fair-NE's balance scores outperformed FairWalk's as we increased the number of clusters, with the gap between them also increasing.

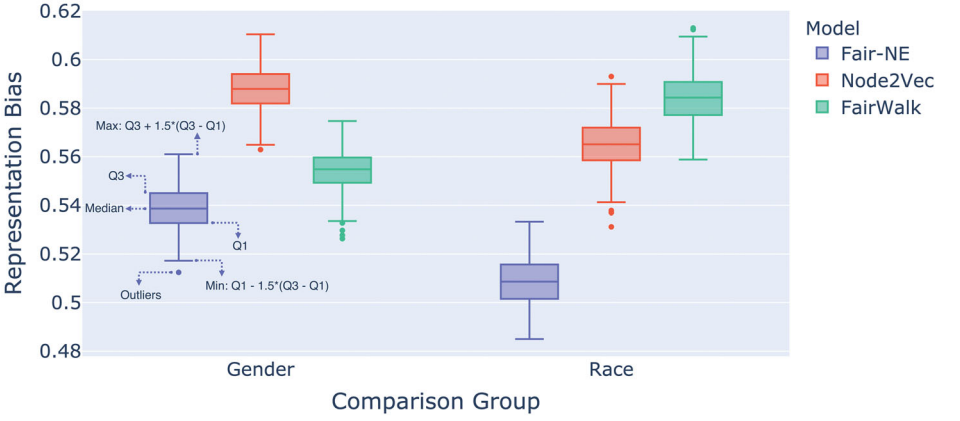


Figure 7. Representation bias of network embeddings in terms of gender and race. *Note.* Values closer to 0.5 are fairer.

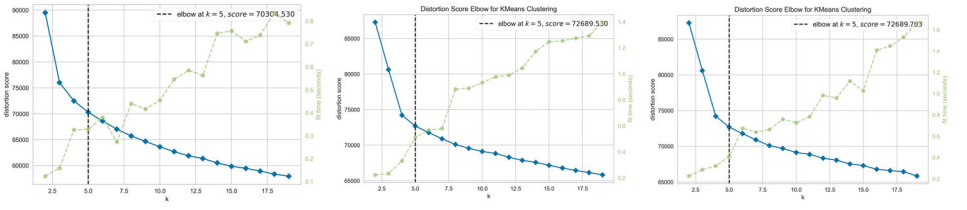


Figure 8. Elbow method results showing best numbers of clusters for network embeddings.

Predictive accuracy of link prediction models

Table 2 shows the AUC and accuracy of the link prediction models using the latent vectors of Fair-NE, Node2Vec, and FairWalk. All three models' high values of predictive accuracy showed that network embedding models could help predict students' probability of connections accurately. Although Fair-NE did not achieve the best performance regarding AUC and accuracy, Fair-NE achieved comparable predictive accuracy than its benchmarks.

Predictive fairness of link prediction models

Figure 10 shows the distributions of EO of link prediction models using different predictive thresholds. Within the threshold range of 0.4 and above, Fair-NE achieved lower or comparable EO than FairWalk. The fairness advantage of Fair-NE was

Table 2. Predictive accuracy of link prediction models.

Models	AUC	Accuracy
Fair-NE	0.9821	0.9326
Node2Vec	0.9866*	0.9480*
FairWalk	0.9464	0.9450

Note. * denotes the best performance of a metric. Greater values indicate better predictive accuracy.

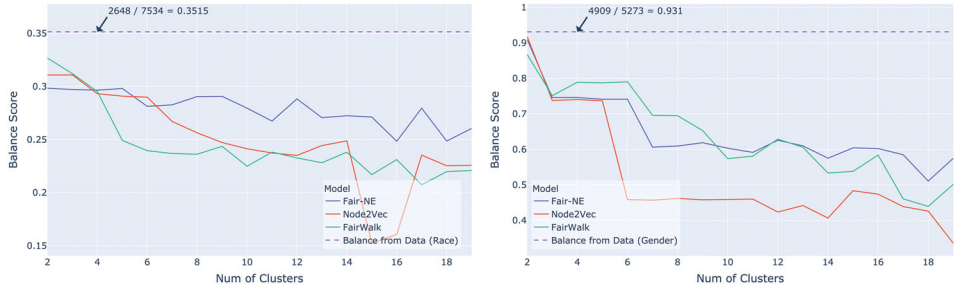


Figure 9. Clustering balance scores of gender and race. *Note.* Values closer to the dotted purple line, Balance from Data, suggest fairer representation.

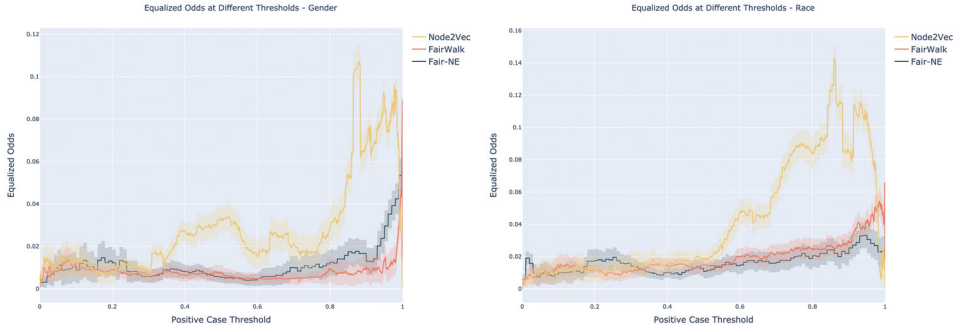


Figure 10. Equalized odds of link prediction models of gender and race at different thresholds. *Note.* Lower values are fairer.

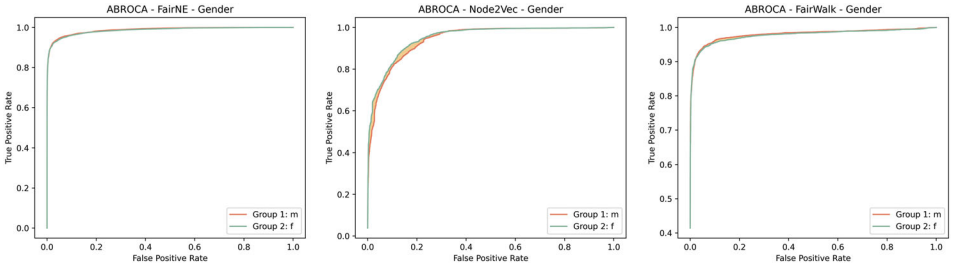


Figure 11. ABROCA of link prediction models of gender. *Note.* Lower values are fairer.

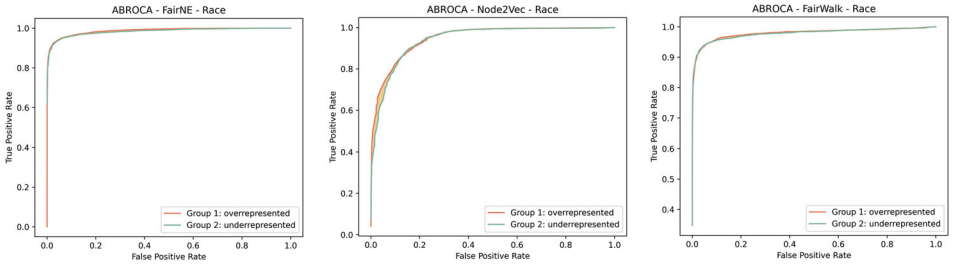


Figure 12. ABROCA of link prediction models of race. *Note.* Lower values are fairer.

apparent when compared with fairness-unaware Node2Vec. Figure 11 illustrates the ABROCAs of link prediction models using gender as the comparison group, and Figure 12 shows those when comparing race. The results showed that Fair-NE achieved the smallest value of ABROCA when comparing gender, indicating the fairest performance. However, FairWalk achieved the fairest result regarding race, with Fair-NE still showing comparable results.

Discussion

Educational studies on AI-based peer recommenders have focused mainly on improving predictive accuracy (e.g., Hansen et al., 2019; Sunar et al., 2017; Xiao et al., 2021), while little has been done to examine the potential fairness issues of those systems. Recent educational studies on AI have shown a burgeoning interest in creating accountable and trustworthy intelligent learning systems. Some studies have conceptualized AI fairness in education by operationalizing the definition of AI fairness in education using notions of equality and equity (Kizilcec & Lee, 2020), introducing metrics to quantify AI fairness in education (Kizilcec & Lee, 2020), and discussing the origins of AI bias as well as their potential harms in education (Baker & Hawn, 2021). Other studies have empirically evaluated fairness of AI models applied in education such as automatic categorization of online discussion posts (Sha et al., 2021), identification of academically at-risk students (Riazy & Simbeck, 2019), and prediction of graduation time (Hutt et al., 2019). Although AI bias was not always present in these evaluations, the researchers stressed the

importance of catering to AI fairness consciously. Educational studies on fair AI have laid a solid foundation to help researchers conceptually understand AI fairness. However, few have developed concrete strategies to proactively address AI bias in education. To the best of our knowledge, this study is the first endeavor to provide actionable strategies to mitigate the AI bias of peer recommenders in education.

Advantages of Fair-NE

We developed Fair-NE to provide recommendations for online peers to support help-seeking. Fair-NE has the following advantages compared to the models evaluated in this study.

First, Fair-NE was designed to be fairness-aware to address bias consciously. Although FairWalk is also fairness-aware, it allows only one protected attribute, meaning multiple FairWalk models are needed if both gender and race are protected (Rahman et al., 2019).

Second, the Bayesian paradigm of Fair-NE indicates that the model can be updated dynamically (König & van de Schoot, 2018). For example, when conducting research with new platforms, researchers or developers of educational peer recommenders might start with small-sized data. Using Fair-NE, researchers can train with a limited amount of data and then incrementally update the model's knowledge with new data effectively and efficiently. On the contrary, Node2Vec and FairWalk need to experience a complete iteration of model training every time new data is available, which can be resource-intensive and time-consuming.

Third, the Bayesian property of Fair-NE allows its users to apply personal insights into the modeling process (Levy, 2016; Xing et al., 2021b). For example, training data to build peer recommenders might not be representative of the distributions of students' demographics. In this case, researchers can incorporate their sources of information (e.g., meta-analysis, governmental reports, and prior experience of platform experts) in Fair-NE's learning. In contrast, the other two models cannot achieve this. Fair-NE's ability to update knowledge incrementally and incorporate user-defined information aligns with the construction of knowledge (Harel & Papert, 1991), whereas Node2Vec and FairWalk adopt a paradigm that is based on the idea of punishment from suboptimized predictive performance (Loftus & Madden, 2020).

Finally, Fair-NE adopts robust negative sampling to allow the model to understand better why students in a dataset are not connected, which can enhance its predictive accuracy (Armandpour et al., 2019). Studies have suggested that there can be trade-offs between model accuracy and fairness (e.g., accuracy might need to be sacrificed for fairness). However, such trade-offs can be mitigated with a careful model design that efficiently utilizes available information (Islam et al., 2021). Table 3 organizes the information on the comparison of Fair-NE, Node2Vec, and FairWalk.

Garbage in, garbage out

Garbage in, garbage out is widely seen across domains such as instructional design (e.g., Snelbecker, 2018) and computer science (e.g., Vidgen & Derczynski, 2020). The

Table 3. Comparison of Fair-NE, Node2Vec, and FairWalk.

Features	Models		
	Fair-NE	Node2Vec	FairWalk
Fairness-aware	√		√
Multiple protected attributes	√		
Incremental update	√		
Incorporation of personal beliefs	√		
Enhanced mechanism for predictive accuracy	√	√	

principle suggests that the quality of output is heavily influenced by that of input data, which also applies to the fairness of AI models in education. The results of descriptive statistics showed that most of the interactions in our dataset happened between students of the same demographic background (135,599 out of 187,450). Inter-gender ($n = 36{,}391$) and inter-racial ($n = 25{,}980$) interactions each took less than 20% of the total number of interactions. The visualization of network communities of students further suggests that students can form communities that are not inclusive enough. Our hypothesis is supported by the fairness evaluation that such a dataset would lead to bias in network embedding models. Results of fairness-unaware model Node2Vec are consistently less fair regarding RB, BS, EO, and ABROCA. The finding that data bias can be embedded in AI models is not news. Baker and Hawn (2021) discussed that the lack of representation of certain demographic groups in data could lead to AI bias against them, which has been found in educational studies. For example, studies have shown that AI could be biased against students with disabilities (Riazy & Simbeck, 2019) or students whose first language is not English in a context where English is used for communication (Sha et al., 2021). In our study, the noninclusive interactions among students can make fairness-unaware model Node2Vec hold the false causal assumption that connections between students with the same demographic factors should be prioritized. Fairness-aware models such as Fair-NE and FairWalk, on the contrary, are designed to better identify the correlation between students' demographics and connection establishment, instead of mistakenly treating correlation as causation.

Experiment of peer recommenders

The results of our experiment indicate that making AI models fairness-aware with enhanced model learning mechanisms of prediction is essential to reducing the effects of data bias while maintaining competitive accuracy. The results, to a great extent, answered our research question of the affordances of fairness-aware network embedding models on prediction fairness and accuracy. The low representation bias of Fair-NE suggests that it can encode students' information about gender and race "agnostically." Educational AI models such as peer recommenders built upon fair latent vectors can learn to make predictions on students without being heavily affected by students' demographic factors (Bose & Hamilton, 2019). The better alignment of Fair-NE's balance scores suggests that clustering students with Fair-NE's latent vectors can result in groups that are fairly represented by demographic attributes. Interestingly, the best number of clusters based on partition ability does not necessarily align with

the numbers that yield good balance scores. This finding resonates with the findings of Quy et al.'s (2021) study. They examined the fairness of clustering algorithms with educational datasets and found inconsistent indications on best models between accuracy and fairness. Moreover, the low EO and ABROCA scores of Fair-NE demonstrate that link prediction models based on Fair-NE tend not to under- or overestimate students of a specific gender or race regarding the probability of establishing connections with a student (Gardner et al., 2019; Kizilcec & Lee, 2020). Finally, Fair-NE can achieve competitive predictive accuracy compared with Node2Vec ($AUC_{Fair\text{-}NE} = 0.982$ vs. $AUC_{Node2Vec} = 0.986$) and better accuracy compared with FairWalk ($AUC_{Fair\text{-}NE} = 0.982$ vs. $AUC_{FairWalk} = 0.946$), suggesting the trade-off between fairness and accuracy can be well balanced in Fair-NE.

Implications for practice

There are three main implications of this study for practitioners and researchers implementing AI-based peer recommenders to support online learning.

First, prepare and scrutinize datasets used in building peer recommenders with group membership to better understand potential data bias. Educational studies on AI have carefully prepared datasets from the perspective of statistical power using methods such as power analysis and missing data imputation (e.g., Larmuseau et al., 2020; Li & Xing, 2021), with few studies putting enough weight on using students' group membership such as demographics for analysis (Paquette et al., 2020). In this study, we were able to identify potential issues of noninclusive communities among students and used the information to yield possible explanations on the bias presented in fairness-unaware model Node2Vec. Similar strategies have been successfully adopted to understand datasets used for automated essay scoring (Loukina & Buzick, 2017) and academic success prediction (Yu et al., 2020). Therefore, collecting data on students' group membership and analyzing datasets with such information are essential to uncover data bias and understand where bias occurs (Baker & Hawn, 2021).

Second, adopt a methodological shift in AI in education from solely focusing on predictive accuracy to stressing the importance of both fairness and accuracy. In this study, we synthesized a variety of fairness metrics and developed a new algorithmic strategy to evaluate and mitigate AI bias of peer recommenders. The desirably fair and comparably accurate performance of Fair-NE suggests that there can be a methodological shift in AI in education, where more efforts are put into building accountable and trustworthy AI systems. As pointed out by Pedro et al. (2019), while AI is transformative in education, it can also be disruptive in that biased AI can favor existing privileged students. Such a standpoint is also held by Vincent-Lancrin and Van der Vlies (2020). They argued that AI bias is a great threat to AI's sustainability in education, potentially further dividing society. Furthermore, addressing AI fairness is the prerequisite for trust-building between students and AI. Tsai et al. (2020) interviewed students on their opinions of AI-based educational systems. Their results showed that the opaqueness of AI fairness is one of the biggest reasons students distrusted AI. Therefore, educational researchers should adopt the techniques of fair AI in their exploration and application with AI systems.

Third, keep scalability in mind when adopting or developing AI models to build peer recommenders. Results in this study suggest that existing fairness-aware models such as FairWalk can contribute to retaining fairness effectively. However, it might not be as scalable as Fair-NE. In large online contexts such as MOOCs with hundreds of thousands of students, the need to train multiple FairWalk models to cater to different sensitive attributes can be unbearable, especially as the size of data and number of protected attributes increase. Moreover, the lack of incremental update of FairWalk makes it challenging to perform cold start (e.g., the application of models with a small data size), thus making it unwieldy at the beginning of course offering (Xing et al., 2021a). On the other hand, Fair-NE allows researchers to incorporate information from previous course offerings to gradually finetune the model as the course proceeds. Therefore, it is advisable to evaluate AI models' scalability before implementing applications such as peer recommenders for actual use.

Limitations

There are several limitations in this study. First, the fairness of the peer recommenders was evaluated mathematically. However, equity in AI for education is more than satisfying fairness metrics (Kizilcec & Lee, 2020). For example, an AI model might yield similar or equal predictive correction and false alarm rates for students with different group memberships. However, it is possible that the knowledge gap between advanced and academically at-risk students can be widened after using AI systems. The effectiveness of AI systems to support students' learning can be associated with students' self-regulation skills and self-efficacy, which students in need of academic help often lack (Walker & Graham, 2021; Yokoyama, 2019). Having tools to address AI is the first step to transit from fairness to equity in AI for education (Baker & Hawn, 2021), this study developed concrete strategies that researchers can adopt to evaluate and enhance fairness of AI-based peer recommenders. Future studies could utilize fairness-aware models to examine their effect on supporting students' learning to understand the equity affordances of AI. Second, this study focused on building peer recommenders with link prediction as an initial step. To build more complex peer recommenders, strategies to address AI fairness in user-modeling approaches are needed. The use of user-modeling can introduce individual differences through the analysis of rich-format data such as trace data. Future studies could consider individual fairness when constructing fair peer recommenders using a user-modeling approach. Individual fairness requires that students with similar backgrounds should have similar predictions from AI models (Kizilcec & Lee, 2020). Finally, we included only data of students in Algebra I. However, Algebra Nation also offers courses such as Geometry (University of Florida, n.d.), where learning contexts and discussion results can differ. Future studies could examine Fair-NE's generalizability with data from diverse course contexts.

Conclusions

Help-seeking is a valuable practice in online discussion forums. This study created a network embedding approach towards building a fair peer recommender to support help-

seeking in online learning. The findings suggest that learners can be supported with help-seeking suggestions both accurately and fairly in online discussion forums at a large scale. The study also enriches the literature on fair AI by extending conceptual discussion and evaluation studies of AI fairness in education and developing concrete algorithmic strategies to improve fairness in building peer recommenders. We have several directions for future research. First, AI fairness could also be evaluated to reveal students' individual opinions on the trustworthiness and accountability of AI (Marcinkowski et al., 2020). Currently, there is not enough evidence revealing the relationship between model fairness and perceived fairness. We plan to integrate Fair-NE, Node2Vec, and FairWalk on Algebra Nation to examine whether students' perceived fairness on AI is affected by algorithmic bias. We also plan to collect empirical data to understand whether fairness-aware models can effectively support students' learning compared with fairness-unaware models, which tend to have better predictive accuracy. Second, statistical tests on fairness metrics are currently missing in the literature. It is challenging for researchers to determine whether one model's fairness is significantly different from another. We plan to develop a statistical test to inform researchers whether significant differences exist among AI models. Third, AI models' transparency is as equally important as its fairness. We will adopt explainable AI techniques to help researchers and students open the black box of AI-based peer recommenders and understand the role of AI explainability on contributing to human-AI collaboration.

Disclosure statement

No potential conflict of interest was declared by the authors.

Funding information

The research reported here was supported by the Institute of Education Sciences, US Department of Education, through Grant R305C160004 to the University of Florida, the University of Florida AI Catalyst Grant -P0195022, and the University of Florida Informatics Institute Seed Funding. The opinions expressed are those of the authors and do not represent the views of the University of Florida, Institute of Education Sciences, or those of the US Department of Education.

Notes on contributors

Chenglu Li is a doctoral student in the Educational Technology Program at the University of Florida. His research interests include learning analytics, Bayesian statistics, fair AI for educational use, educational software development, and educational games.

Wanli Xing is an assistant professor of educational technology at the University of Florida. His research interests are artificial intelligence, learning analytics, STEM education, and online learning.

Walter L. Leite is a full professor of research and evaluation methodology at the University of Florida. His research interests are exploring how data mining and machine learning methods may assist in statistical modeling for theory development and causal inference.

ORCID

Chenglu Li http://orcid.org/0000-0002-1782-0457
Wanli Xing http://orcid.org/0000-0002-1446-889X
Walter L. Leite http://orcid.org/0000-0001-7655-5668

Data availability statement

The data that support the findings of this study are available from the corresponding author, WX, upon reasonable request.

References

Abu-Salih, B., Al-Tawil, M., Aljarah, I., Faris, H., Wongthongtham, P., Chan, K. Y., & Beheshti, A. (2021). Relational learning analysis of social politics using knowledge graph embedding. *Data Mining and Knowledge Discovery*, *35*, 1497–1536. https://doi.org/10.1007/s10618-021-00760-w

Allen, K.-A., & Kern, M. L. (2017). *School belonging in adolescents: Theory, research and practice*. Springer. https://doi.org/10.1007/978-981-10-5996-4

Armandpour, M., Ding, P., Huang, J., & Hu, X. (2019). Robust negative sampling for network embedding. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(1), 3191–3198. https://doi.org/10.1609/aaai.v33i01.33013191

Baker, R. S., & Hawn, A. (2021). *Algorithmic bias in education*. EdArXiv. https://doi.org/10.35542/osf.io/pbmvz

Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, *20*(1), 1–13. https://doi.org/10.1080/1369118X.2016.1216147

Bose, A. J., & Hamilton, W. L. (2019). Compositional fairness constraints for graph embeddings. *Proceedings of Machine Learning Research*, *97*, 715–724. http://proceedings.mlr.press/v97/bose19a/bose19a.pdf

Bouchet, F., Labarthe, H., Yacef, K., & Bachelet, R. (2017). Comparing peer recommendation strategies in a MOOC. In M. Tkalcic, D. Thakker, P. Germanakos, K. Yacef, C. Paris, & O. Santos (Eds.), *UMAP '17: Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (pp. 129–134). Association for Computing Machinery. https://doi.org/10.1145/3099023.3099036

Buyl, M., & De Bie, T. (2020). DeBayes: A Bayesian method for debiasing network embeddings. *Proceedings of Machine Learning Research*, *119*, 1220–1229. http://proceedings.mlr.press/v119/buyl20a/buyl20a.pdf

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Chai, Y., Cong, Y., Sun, R., Wu, F., Zhang, Z., Wan, Y., & Cui, L. (2019). Finding potential empathizers in an online mental health community: A deep graph embedding approach. In *Proceedings of 2019 IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking* (pp. 1646–1651). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ispa-bdcloud-sustaincom-socialcom48970.2019.00243

Chao, P.-Y., Lai, K. R., Liu, C.-C., & Lin, H.-M. (2018). Strengthening social networks in online discussion forums to facilitate help seeking for solving problems. *Educational Technology & Society*, *21*(4), 39–50. https://drive.google.com/file/d/17ZV_vFFY9ZHOzrW_ptZG_1Fm7Zs975WP/view

Chierichetti, F., Kumar, R., Lattanzi, S., & Vassilvitskii, S. (2017). Fair clustering through fairlets. In U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, & R. Fergus (Eds.), *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 5036–5044). Association for Computing Machinery. https://dl.acm.org/doi/abs/10.5555/3295222.3295256

Chyr, W.-L., Shen, P.-D., Chiang, Y.-C., Lin, J.-B., & Tsai, C.-W. (2017). Exploring the effects of online academic help-seeking and flipped learning on improving students' learning. *Educational Technology & Society*, *20*(3), 11–23. https://drive.google.com/file/d/12wjgOwrefvjQmJwQcoKr8RDKvv08wYYh/view

Costa, P. T., Jr, & McCrae, R. R. (2008). The revised NEO personality inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessmen: Vol. 2. Personality measurement and testing* (pp. 179–198). Sage Publications, Inc. https://doi.org/10.4135/9781849200479.n9

Cross, S., Waters, Z., Kitto, K., & Zuccon, G. (2017). Classifying help seeking behaviour in online communities. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference (pp. 419–423). Association for Computing Machinery. https://doi.org/10.1145/3027385.3027442

Esposito, C., Landrum, G. A., Schneider, N., Stiefl, N., & Riniker, S. (2021). GHOST: Adjusting the decision threshold to handle imbalanced data in machine learning. *Journal of Chemical Information and Modeling*, *61*(6), 2623–2640. https://doi.org/10.1021/acs.jcim.1c00160

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. https://doi.org/10.1038/nature21056

Garcia-Martinez, S., & Hamou-Lhadj, A. (2013). Educational recommender systems: A pedagogical-focused perspective. In G. A. Tsihrintzis, M. Virvou, & L. C. Jain (Eds.), *Multimedia services in intelligent environments: Recommendation services* (pp. 113–124). Springer. https://doi.org/10.1007/978-3-319-00375-7_8

Gardner, J., Brooks, C., & Baker, R. (2019). Evaluating the fairness of predictive student models through slicing analysis. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge (pp. 225–234). Association for Computing Machinery. https://doi.org/10.1145/3303772.3303791

Garg, M., & Goel, A. (2021). A data-driven approach for peer recommendation to reduce dropouts in MOOC. In S. M. Thampi, E. Gelenbe, M. Atiquzzaman, V. Chaudhary, & K. C. Li (Eds.), *Lecture notes in electrical engineering: Vol. 735. Advances in computing and network communications* (pp. 217–229). Springer. https://doi.org/10.1007/978-981-33-6977-1_18

Grover, A., & Leskovec, J. (2016). Node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 855–864). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939754

Hansen, P., Bustamante, R. J., Yang, T.-Y., Tenorio, E., Brinton, C., Chiang, M., & Lan, A. (2019). Predicting the timing and quality of responses in online discussion forums. In *Proceedings of the IEEE 39th International Conference on Distributed Computing Systems* (pp. 1931–1940). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ICDCS.2019.00191

Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*, *29*, 3315–3323. https://proceedings.neurips.cc/paper/2016/file/9d2682367c3935defcb1f9e247a97c0d-Paper.pdf

Hutt, S., Gardner, M., Duckworth, A. L., & D'Mello, S. K. (2019). Evaluating fairness and generalizability in models predicting on-time graduation from college applications. In C. Lynch, A. Merceron, M. Desmarais, & R. Nkambou (Eds.), *Proceedings of the 12th International Conference on Educational Data Mining* (pp. 79–88). International Educational Data Mining Society. https://drive.google.com/file/d/1O2CEzb09h1kon2wEyuVJV9mCWDKDjPWf/view

Islam, R., Pan, S., & Foulds, J. R. (2021). Can we obtain fairness for free? In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (pp. 586–596). Association for Computing Machinery. https://doi.org/10.1145/3461702.3462614

Kear, K. (2004). Peer learning using asynchronous discussion systems in distance education. *Open Learning: The Journal of Open, Distance and e-Learning*, *19*(2), 151–164. https://doi.org/10.1080/0268051042000224752

Kizilcec, R. F., & Lee, H. (2020). *Algorithmic fairness in education*. ArXiv. https://arxiv.org/pdf/2007.05443.pdf

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of cluster in k-means clustering. *International Journal of Advance Research in Computer Science and Management Studies*, *1*(6), 90–95. http://www.ijarcsms.com/docs/paper/volume1/issue6/V1I6-0015.pdf

König, C., & van de Schoot, R. (2018). Bayesian statistics in educational research: A look at the current state of affairs. *Educational Review*, *70*(4), 486–509. https://doi.org/10.1080/00131911.2017.1350636

Labarthe, H., Bouchet, F., Bachelet, R., & Yacef, K. (2016). Does a peer recommender foster students' engagement in MOOCs? In T. Barnes, M. Chi, & M. Feng (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 418–423). International Educational Data Mining Society. https://www.educationaldatamining.org/EDM2016/proceedings/paper_171.pdf

Larmuseau, C., Cornelis, J., Lancieri, L., Desmet, P., & Depaepe, F. (2020). Multimodal learning analytics to investigate cognitive load during online problem solving. *British Journal of Educational Technology*, *51*(5), 1548–1562. https://doi.org/10.1111/bjet.12958

Levy, R. (2016). Advances in Bayesian modeling in educational research. *Educational Psychologist*, *51*(3–4), 368–380. https://doi.org/10.1080/00461520.2016.1207540

Li, C., & Xing, W. (2021). Natural language generation using deep learning to support MOOC learners. *International Journal of Artificial Intelligence in Education*, *31*, 186–214. https://doi.org/10.1007/s40593-020-00235-x

Li, C., Xing, W., & Leite, W. (2021). Yet another predictive model? Fair predictions of students' learning outcomes in an online math learning platform. In Proceedings of the 11th International Learning Analytics and Knowledge Conference (pp. 572–578). Association for Computing Machinery. https://doi.org/10.1145/3448139.3448200

Loftus, M., & Madden, M. G. (2020). A pedagogy of data and artificial intelligence for student subjectification. *Teaching in Higher Education*, *25*(4), 456–475. https://doi.org/10.1080/13562517.2020.1748593

Loukina, A., & Buzick, H. (2017). Use of automated scoring in spoken language assessments for test takers with speech impairments. *ETS Research Report Series*, *2017*(1), 1–10. https://doi.org/10.1002/ets2.12170

Marcinkowski, F., Kieslich, K., Starke, C., & Lünich, M. (2020). Implications of AI (un-) fairness in higher education admissions: The effects of perceived AI (un-) fairness on exit, voice and organizational reputation. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 122–130). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372867

Melrose, S. (2006). Facilitating help-seeking through student interactions in a WebCT online graduate study program. *Nursing & Health Sciences*, *8*(3), 175–178. https://doi.org/10.1111/j.1442-2018.2006.00277.x

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Proceedings of the 26th International Conference on Neural Information Processing Systems* (vol. 2, pp. 3111–3119). Curran Associates Inc. https://dl.acm.org/doi/10.5555/2999792.2999959

National Science Foundation. (2017). *Women, minorities, and persons with disabilities in science and engineering: 2017* (Special Report NSF 17-310). https://www.nsf.gov/statistics/2017/nsf17310/digest/about-this-report/

Nelson, W., Zitnik, M., Wang, B., Leskovec, J., Goldenberg, A., & Sharan, R. (2019). To embed or not: Network embedding as a paradigm in computational biology. *Frontiers in Genetics*, *10*, Article 381. https://doi.org/10.3389/fgene.2019.00381

Harel, I. E., & Papert, S. E. (1991). *Constructionism*. Ablex Publishing Corporation.

Office for Civil Rights. (2016). *College and career readiness*. https://ocrdata.ed.gov/estimations/2015-2016

Paquette, L., Ocumpaugh, J., Li, Z., Andres, A., & Baker, R. (2020). Who's learning? Using demographics in EDM research. *Journal of Educational Data Mining*, *12*(3), 1–30. https://jedm.educationaldatamining.org/index.php/JEDM/article/view/404

Parnes, M. F., Kanchewa, S. S., Marks, A. K., & Schwartz, S. E. (2020). Closing the college achievement gap: Impacts and processes of a help-seeking intervention. *Journal of Applied Developmental Psychology*, *67*, Article 101121. https://doi.org/10.1016/j.appdev.2020.101121

Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). *Artificial intelligence in education: Challenges and opportunities for sustainable development* (Report No. ED-2019/WS/8). United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000366994

Phillips, T., & Ozogul, G. (2020). Learning analytics research in relation to educational technology: Capturing learning analytics contributions with bibliometric analysis. *TechTrends*, *64*, 878–886. https://doi.org/10.1007/s11528-020-00519-y

Potts, B. A., Khosravi, H., Reidsema, C., Bakharia, A., Belonogoff, M., & Fleming, M. (2018). Reciprocal peer recommendation for learning purposes. In Proceedings of the 8th International Conference on Learning Analytics and Knowledge (pp. 226–235). Association for Computing Machinery. https://doi.org/10.1145/3170358.3170400

Quy, T. L., Roy, A., Friege, G., & Ntoutsi, E. (2021). *Fair-capacitated clustering*. ArXiv. https://arxiv.org/pdf/2104.12116.pdf

Rahman, T. A., Surma, B., Backes, M., & Zhang, Y. (2019). Fairwalk: Towards fair graph embedding. In S. Kraus (Ed.), *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence* (pp. 3289–3295). International Joint Conferences on Artificial Intelligence. https://doi.org/10.24963/ijcai.2019/456

Riazy, S., & Simbeck, K. (2019). Predictive algorithms in learning analytics and their fairness. In N. Pinkwart & J. Konert (Eds.), *Proceedings of DELFI 2019* (pp. 223–228). Gesellschaft für Informatik eV. https://doi.org/10.18420/delfi2019_305

Sha, L., Rakovic, M., Whitelock-Wainwright, A., Carroll, D., Yew, V. M., Gasevic, D., & Chen, G. (2021). Assessing algorithmic fairness in automatic classifiers of educational forum posts. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova (Eds.), *Lecture notes in computer science: Vol. 12748. Artificial intelligence in education* (pp. 381–394). Springer. https://doi.org/10.1007/978-3-030-78292-4_31

Shaw, E. (2019). *A study of factors and perceptions that mediate student participation in supplementary discussion forums* (Publication No. 27926111) [Doctoral dissertation, Open University]. ProQuest Dissertations & Theses Global. https://www.proquest.com/docview/2375928549

Smith, P. S., Nelson, M. M., Trygstad, P. J., & Banilower, E. R. (2013). *Unequal distribution of resources for K-12 science instruction: Data from the 2012 National Survey of Science and Mathematics Education*. Horizon Research. http://www.horizon-research.com/2012nssme/wp-content/uploads/2013/06/NARST-2013-Equity-paper-revised-and-final.pdf

Snelbecker, G. E. (2018). Contrasting and complementary approaches to instructional design. In C. Reigeluth (Ed.), *Instructional theories in action* (pp. 321–337). Routledge. https://doi.org/10.4324/9780203056783-10

Sun, J., Geng, J., Cheng, X., Zhu, M., Xu, Q., & Liu, Y. (2020). Leveraging personality information to improve community recommendation in e-learning platforms. *British Journal of Educational Technology*, *51*(5), 1711–1733. https://doi.org/10.1111/bjet.13011

Sunar, A. S., White, S., Abdullah, N. A., & Davis, H. C. (2017). How learners' interactions sustain engagement: A MOOC case study. *IEEE Transactions on Learning Technologies*, *10*(4), 475–487. https://doi.org/10.1109/TLT.2016.2633268

Tsai, Y.-S., Perrotta, C., & Gašević, D. (2020). Empowering learners with personalised learning approaches? Agency, equity and transparency in the context of learning analytics. *Assessment & Evaluation in Higher Education*, *45*(4), 554–567. https://doi.org/10.1080/02602938.2019.1676396

University of Florida. (n.d.). *Mathematics*. Retrieved November 4, 2021, from https://lastinger.center.ufl.edu/mathematics/

van den Broek, E., Sergeeva, A., & Huysman, M. (2020). Hiring algorithms: An ethnography of fairness in practice. In *Proceedings of the 40th International Conference on Information Systems* (pp. 1–9). Association for Information Systems. https://aisel.aisnet.org/icis2019/future_of_work/future_work/6/

Vidgen, B., & Derczynski, L. (2020). Directions in abusive language training data, a systematic review: Garbage in, garbage out. *Plos one*, *15*(2), Article e0243300. https://doi.org/10.1371/journal.pone.0243300

Vincent-Lancrin, S., & Van der Vlies, R. (2020). Trustworthy artificial intelligence (AI) in education: Promises and challenges. *OECD Education Working Papers*, *218*, 1–13. https://doi.org/10.1787/a6c90fa9-en

Walker, S., & Graham, L. (2021). At risk students and teacher-student relationships: Student characteristics, attitudes to school and classroom climate. *International Journal of Inclusive Education*, *25*(8), 896–913. https://doi.org/10.1080/13603116.2019.1588925

Wu, L., Zhang, Q., Chen, C.-H., Guo, K., & Wang, D. (2020). Deep learning techniques for community detection in social networks. *IEEE Access*, *8*, 96016–96026. https://doi.org/10.1109/ACCESS.2020.2996001

Xiao, X., Sun, R., Yao, Z., Zhang, C., & Chen, X. (2021). A novel framework with weighted heterogeneous educational network embedding for personalized freshmen recommendation under the impact of COVID-19 storm. *IEEE Access*, *9*, 67129–67142. https://doi.org/10.1109/ACCESS.2021.3075675

Xing, W., Du, D., Bakhshi, A., Chiu, k. C., & Du, H. (2021a). Designing a transferable predictive model for online learning using a Bayesian updating approach. *IEEE Transactions on Learning Technologies*, *14*, 474–485. https://doi.org/10.1109/TLT.2021.3107349

Xing, W., Li, C., Chen, G., Huang, X., Chao, J., Massicotte, J., & Xie, C. (2021b). Automatic assessment of students' engineering design performance using a Bayesian network model. *Journal of Educational Computing Research*, *59*, 230–256. https://doi.org/10.1177/0735633120960422

Xu, B., & Yang, D. (2015). Study partners recommendation for xMOOCs learners. *Computational Intelligence and Neuroscience*, *15*, 1–10. https://doi.org/10.1155/2015/832093

Xu, Z., Ou, Z., Su, Q., Yu, J., Quan, X., & Lin, Z. (2020). Embedding dynamic attributed networks by modeling the evolution processes. In D. Scott, N. Bel, & C. Zong (Eds.), *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 6809–6819). International Committee on Computational Linguistics. https://aclanthology.org/2020.coling-main.600.pdf https://doi.org/10.18653/v1/2020.coling-main.600

Yang, T.-Y., Brinton, C. G., & Joe-Wong, C. (2018). Predicting learner interactions in social learning networks. In *IEEE INFOCOM 2018—Proceedings of the IEEE Conference on Computer Communications* (pp. 1322–1330). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/INFOCOM.2018.8485927

Yokoyama, S. (2019). Academic self-efficacy and academic performance in online learning: A mini review. *Frontiers in psychology*, *9*, Article 2794. https://doi.org/10.3389/fpsyg.2018.02794

Yu, R., Li, Q., Fischer, C., Doroudi, S., & Xu, D. (2020). Towards accurate and fair prediction of college success: Evaluating different sources of student data. In A. Rafferty, J. Whitehill, V. Cavalli-Sforza, & C. Romero (Eds.), *Proceedings of the 13th International Conference on Educational Data Mining* (pp. 292–301). International Educational Data Mining Society. https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_194.pdf

Zemel, R., Wu, Y., Swersky, K., Pitassi, T., & Dwork, C. (2013). *Proceedings of Machine Learning Research*, *28*, 325–333. https://proceedings.mlr.press/v28/zemel13.html

Zimmer, F., Scheibe, K., Stock, M., & Stock, W. G. (2019). Fake news in social media: Bad algorithms or biased users? *Journal of Information Science Theory and Practice*, *7*(2), 40–53. https://doi.org/10.1633/JISTAP.2019.7.2.4