Research Article

An Advanced Deep Attention Collaborative Mechanism for Secure Educational Email Services

Yanfang Chen and Yongzhao Yang

Zhengzhou Preschool Education College, Zhengzhou 450000, China

Correspondence should be addressed to Yanfang Chen; chenyanfang0822@163.com

Received 3 March 2022; Revised 18 March 2022; Accepted 12 April 2022; Published 26 April 2022

Academic Editor: Konstantinos Demertzis

Copyright © 2022 Yanfang Chen and Yongzhao Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The COVID-19 crisis has once again highlighted the vulnerabilities of some critical areas in cyberspace, especially in the field of education, as distance learning and social distance have increased their dependence on digital technologies and connectivity. Many recent cyberattacks on e-learning systems, educational content services, and trainee management systems have created severe demands for specialized technological solutions to protect the security of modern training methods. Email is one of the most critical technologies of educational organizations that are attacked daily by spam, phishing campaigns, and all kinds of malicious programs. Considering the efforts made by the global research community to ensure educational processes, this study presents an advanced deep attention collaborative filter for secure academic email services. It is a specialized application of intelligent techniques that, for the first time, examines and models the problem of spam as a system of graphs where collaborative referral systems undertake the processing and analysis of direct and indirect social information to detect and categorize spam emails. In this study, nonnegative matrix factorization (NMF) is applied to the social graph adjacent table to place users in one (or more) overlapping communities. Also, using a deep attention mechanism, it becomes personalized for each user. At the same time, with the introduction of exponential random graph models (ERGMs) in the process of factorization, local dependencies are significantly mitigated to achieve the revelation of malicious communities. This methodology is being tested successfully in implementing mail protection systems for educational organizations. According to the findings, the proposed algorithm outperforms all other compared algorithms in every metric tested.

1. Introduction

During the coronavirus pandemic, the rapid and violent digital transformation is called upon to deal with a massive wave of sophisticated and persistent cyberattacks related to the new reality [1–4]. This new reality is becoming particularly evident in education, which has become one of the ideal targets for digital attacks, as distance learning has become a necessity for billions of students worldwide [5].

In particular, advanced phishing campaigns are constantly evolving, using e-learning and distance education, access to education services, and educational content management systems as a theme. Schools and universities have switched to large-scale e-learning platforms, often without knowing crucial privacy issues [6]. Students in distance learning programs are also attacked daily. Furthermore, the rate of ransomware attacks is growing exponentially, with specialized cybercriminals first extracting large amounts of sensitive personal data before encrypting the educational databases of an institution or training organization. They then threaten to publish these data unless a ransom is paid, putting additional pressure on the organizations called upon to meet the criminals’ demands in question [7].

The recipients are targeted by phishing attacks, which imitate the login portals of universities to steal credentials, etc. In most cases, these scams are related to the email of educational organizations, where the fundamental security flaws can be easily bypassed. In most phishing campaigns targeting educational organizations, the attack begins with
an email that supposedly contains information about the institutions’ instructions on complying with the COVID-19 protocols set by the relevant ministries, with instructions for course changes, grading, hypothetical links to digital classrooms, etc. Misleading spam emails encourage the recipient to click on an attached HTML file, which leads them to a fake login page similar to the teacher’s login page, with disastrous results. These pages look very convincing, and the URLs use a similar naming pattern that includes the top-level domain of the relevant educational organizations [8, 9].

The most common and widely used machine learning technique for protection against related threats and the intelligent classification of spam emails is based on the use of measures that are based on distance. These methods are considered supervised learning methods, which presuppose that the whole set includes the input data and the desired categorization for each element [10, 11].

To categorize each new element in a class, it is necessary to calculate its distance from each part of the training set, finally considering only the assignments closest to it. These methods are based on analogy and not on producing a generalization model. Thus, there is no training stage, and no model is created until a new observation needs to be categorized. For this reason, the relevant categorizers are also called lazy classifiers [12]. Also, the specific methods for classifying a new observation must compare it with available observations of the training set, which requires storing all or at least part of the training data.

Although these methods can significantly achieve the complexity of detecting complex dependencies between the variables that make up the problem, they are relatively simple to implement and use and can achieve high classification performance [12, 13]. Many comparisons between observations require very effective indexing techniques. Their categorization of new observations takes longer and requires high availability of computing resources. Furthermore, the classification results are sensitive to the local characteristics of the data, the existence of insignificant input variables, and the number of categorization observations, increasing the risk of over-adaptation. We propose an innovative methodology to use a personalized attention mechanism to overcome the relevant obstacles.

The study is organized as follows: Section 2 provides an overview of the various relevant approaches that have been identified in the literature. Section 3 presents the proposed methodology. The scenarios and results are presented in Section 4. Finally, a summary of the findings and a list of potential future research directions are concluded in Section 5.

2. Literature Review

The recent literature concerning the field of detection mechanisms for email services is as follows.

Abdullahi and Kaya [14] suggested a deep learning system for detecting phishing in emails and messages. On both email and SMS collections, they utilized ML algorithms such as SVM classifier, multinomial naive Bayes, decision tree, random forest, logistic regression, and dense neural network. They used existing assessment criteria to evaluate the classifiers on the datasets they employed. The analysis was carried out using coding techniques and the TensorFlow technology, and the outcome revealed that dense neural networks outperform deep learning classifications in identifying phishing attempts in all the samples. The suggested strategy outperformed traditional machine learning algorithms on real datasets.

Fang et al. [15] employed an algorithm dubbed THEMIS to identify suspicious emails. They began by analyzing the email layout and then recommended their scheme, which was used to model emails at the email header, body, character, and word level simultaneously, using an improved recurrent convolutional neural network (RCNN) model with multilevel vectors and attention mechanism. This approach was used in both the header and the model’s body, causing it to pay greater attention to the more critical information between them. They utilized an imbalanced dataset with actual ratios of phising and genuine emails to perform tests and assess THEMIS achieving overall accuracy high levels, according to the testing data. In the meanwhile, the false-positive rate was low. The filter’s accuracy and low FPR help detect phishing emails with high likelihood, while benevolent emails were filtered out as little as possible. They aim to improve their model for identifying phishing emails that do not have a header and simply have content.

Phomkeona and Okamura [16] proposed a method for categorizing and diagnosing zero-day malicious emails based on data gathered from the email header and content. They merged it with dynamic analytic data as a collection of 27 features, including machine translation detection, risk word detection, and other characteristics, using numerous application programming interfaces. To teach and evaluate the system, four distinct language email datasets were employed to replicate real-world diversity and zero-day harmful email attack scenarios. They achieved a reasonable detection rate for both zero-day malicious email types and regular spam. They stated that by adding new contaminated spam datasets to train the algorithm and utilizing a translation API to boost accuracy, their model could be improved.

Kaddoura et al. [17] applied deep learning methods to identify link-less emails, and they proposed a spam email detection system based on FFNN. Different settings were used to optimize hyperparameters. The Enron dataset was preprocessed, and two feature extraction algorithms were used. On the Enron dataset, their model was tested to classify emails as spam or regular. This approach is compared to the BERT dataset. They also looked at numerous variants of its design in terms of layer count, the number of neurons per layer, and the number of neurons per input layer and calculated the F1 score for each one. Precision, recall, and accuracy were calculated to illustrate the approach’s success.

Asudani et al. [18] investigated the efficiency of a pre-trained embedding model for email categorization using deep learning filters such as the long short-term memory and convolutional neural network models. They employed pretrained word embedding using global vectors (GloVe) and Bidirectional Encoder Representations Transformers (BERT) to discover links between words, which helped them...
categorize messages into relevant categories using machine learning models. They concluded that word embedding models boost the accuracy of the email classifier. The experimentation used two benchmark datasets: SpamAssassin and Enron. GloVe embedding achieved quicker execution and improved performance on massive datasets, according to the research. Traditional machine learning methods categorize an email as benevolent or spam, and the CNN to the research. Traditional machine learning methods and Enron. GloVe embedding achieved quicker execution models boost the accuracy of the email classifier. VJ_hey ex-

1. Proposed Methodology

Nonnegative factorization [24] of arrays is part of a broader family of dimensional techniques, which attempt to construct a partial representation of very high-dimensional data by projecting them into a lower-dimensional space. The difference in NMF from other methods is the limitation of the nonnegativity of the elements of the generated tables, which allows a better interpretation of the result. In the specific case we are considering, let \( A \in \mathbb{R}^{m \times n} \) be a table adjacent to a graph of \( n \) nodes. We want to factorize it into two nonnegative arrays: the basis matrix \( W \in \mathbb{R}^{r \times n} \) and the coefficient matrix \( H \in \mathbb{R}^{r \times m} \), so that [25–27]:

\[
A \approx WH,
\]

where \( r \) is the number of communities and \( r \ll n \).

The purpose of the NMF is to calculate the elements of \( W \) and \( H \) so that their product is as “close” as possible to \( A \), with proximity being measured by some distance function. More strictly, nonnegative table factorization is the following combination optimization problem (which in this case is described as a minimization problem) [28, 29]:

\[
\min_{W,H} D(A\|WH),
\]

Subject to, \( W \geq 0, H \geq 0, \)

where \( D(\cdot\|\cdot) \) is a distance function of arrays \( A \) and \( WH \). In general, the problem of minimizing the function \( D \) is NP-hard, for which additionally no convex formulations are known that would lead to finding the total minimum of \( D \) concerning both tables \( W \) and \( H \) at the same time. Although the optimization of \( D \) is non-convex for both arrays simultaneously, it is nevertheless convex for each of the two arrays separately; i.e., keeping unchanged, e.g., \( W \), the problem becomes convex for \( H \) (and vice versa) [30].

To create an immediate and accurate forecasting process, we use the Bayesian NMF and the benefits through the retrospective allocation. The Bayesian NMF is a subcategory of the probabilistic NMF, which approximates the parameters \( W \) and \( H \) using the classical relation of the Bayesian inference [31, 32]:

\[
P(W, H, \Theta|A) \propto P(A|W, H, \Theta) P(W, H|\Theta), P(\Theta),
\]

where the base and coefficient tables are the parameters of the model and \( \Theta \) is the space of the hyperparameters (which regulate the statistical behavior of the distribution from which the tables \( W \) and \( H \) are derived). To use the above relation, a necessary condition is to speculate on the statistical origin of the neighborhood table data and the
product’s table factors. The left part of the relation expresses the a posteriori probability that the model parameters receive a specified value based on the specific data. In contrast, the first term of the product of the right part of the same relation expresses the likelihood of the model, the next is the a priori probability of model parameters for the specific hyperparameters, and finally, the last term is the likelihood of occurrence of the hyperparameters themselves. A careful selection of the probability function and the ex ante probability can result in an algorithm that exhibits better and faster convergence [33, 34]. Thus, the model’s parameters must be observed, and based on this observation, the probability distribution that best expresses their statistical properties must be selected. Then, after the probability formation has been clarified, the appropriate pre-possibility for the area of the hyperparameters is set. To choose the proper a priori probability and probability function, factorization is performed, so the arrangement of the edges between the nodes is expected to show star-type phenomena, i.e., some few nodes with a large number of tangent edges and many nodes with a small number of tangent edges [35–37].

Therefore, there are many “open” triangles, and the most suitable ERGM for the occasion is the 2-star model, which is calculated as [34, 38, 39]:

\[ H = \theta m(G) + ts(G), \]

\[ m(G) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j}, s(G) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} \sum_{k=1}^{n-1} a_{k,j}, \quad (4) \]

where \( a_{ij} \) is the element of the graph adjacent table \( A \) (with value 1 if there is an edge between nodes \( i \) and \( j \), and 0 otherwise), \( m(G) \) is the network statistic that models the number of edges of the graph (whose influence is controlled by the hyperparameter \( \theta \)), and \( s(G) \) is the corresponding magnitude for the number of 2 stars (whose influence is controlled by the hyperparameter \( \tau \), respectively). It should also be noted that in this case, the graph is nondirectional (i.e., for the \( i \) and \( j \) elements of the neighborhood table, the equation \( a_{ij} = a_{ji} \) applies). Since there is no detailed solution for the model described, we must resort to approximate techniques for estimating the value of the hyperparameters \( \theta \) and \( \tau \). As a first step, we rewrite the network statistics as a function, not of the edges, but the degree of each node \( k_i \) each node [33, 39, 40]:

\[ m(G) = \frac{1}{2} \sum_{i=1}^{n} k_i, s(G) = \frac{1}{2} \sum_{i=1}^{n} k_i^2 \times (k_i - 1) = \frac{1}{2} \sum_{i=1}^{n} k_i^2 - m(G). \quad (5) \]

Then, the Hamiltonian is derived as follows:

\[ H = \theta m(G) + ts(G) = \frac{\theta}{2} \sum_{i=1}^{n} k_i^2 + \frac{\theta - \tau}{2} \sum_{i=1}^{n} k_i. \quad (6) \]

To facilitate the following calculations, the hyperparameters \( \theta \) and \( \tau \) are replaced by the auxiliary hyperparameters \( J \) and \( B \), which are defined as follows:

\[ J = \frac{(n-1)\tau}{2}, \quad B = \frac{\tau - \theta}{2}, \quad (7) \]

so the Hamiltonian takes its final form:

\[ H = \frac{J}{n-1} \sum_{i=1}^{n} (k_i)^2 + B \sum_{i=1}^{n} k_i. \quad (8) \]

Comparing the equations, we observe that they express the same ERGM [41], i.e., the 2-star model, using different network statistics. The free energy of the model described by the Hamiltonian is calculated as follows [40, 42, 43]:

\[ F = -n(n-1)J(\phi_0)^2 + \frac{1}{2}n(n-1)\ln(1 + e^{4\phi_0+2B}) \]

\[ + \frac{n}{2} \ln[(n-1)J] - \frac{n}{2} \ln 4\pi, \]

where \( \phi_0 \) is the solution of the mean field approach given as follows:

\[ \phi_0 = \frac{1}{2} \left( \tanh(2J\phi_0 + B) + 1 \right). \quad (10) \]

The partial derivative of the free energy of the model for the hyperparameter \( B \) is equal to the expected value of the sum of the degrees of the nodes:

\[ \langle \sum_{i=1}^{n} k_i \rangle = \frac{\partial F}{\partial B} = \sum_{i=1}^{n} \langle k_i \rangle = \frac{\partial F}{\partial B} \quad (11) \]

By approximating the expected value of the degree of each node \( \langle k_i \rangle \) with its most probable/expected value, i.e., the predicted value \( \langle k_i \rangle \) of the degree of all nodes, we have [44] the following:

\[ \sum_{i=1}^{n} \langle k_i \rangle = \frac{\partial F}{\partial B} \cdot n \langle k \rangle = \frac{\partial F}{\partial B} \langle k \rangle = 1 \frac{\partial F}{n \partial B} \Rightarrow \]

\[ \langle k \rangle = (n-1) \frac{e^{4\phi_0+2B}}{1 + e^{4\phi_0+2B}} = (n-1) \frac{1}{2} \left[ \tanh(2J\phi_0 + B) \right] \]

\[ \langle k \rangle = (n-1)\phi_0 \Rightarrow \phi_0 = \frac{n-1}{n} \langle k \rangle, \quad (12) \]

making \( \phi_0 \) equal to:

\[ \phi_0 = \frac{\langle k \rangle}{n-1}, \quad \phi_0 \in (0,1). \quad (13) \]

Similarly, the partial derivative of free energy for the auxiliary hyperparameter \( J/n-1 \) is equal to the expected value of the sum of the squares of the degrees of the nodes [45]:

\[ \left( \sum_{i=1}^{n} k_i^2 \right) = \frac{\partial F}{\partial (J/n-1)} \]

\[ \sum_{i=1}^{n} \langle k_i^2 \rangle = \frac{n}{\partial (J/n-1)} \]

So, the expected value of the square of the degree of a node \( \langle k_i^2 \rangle \) is approximated by its most probable/expected
value, that is, the expected value of the square of the degree of all nodes [28, 34, 46]:

\[
\langle \sum_{i=1}^{n} k_{i}^{2} \rangle = \sum_{i=1}^{n} \langle k_{i}^{2} \rangle = \frac{\partial F}{\partial (f/n-1)} = n(\langle k^{2} \rangle = (n-1) \frac{\partial F}{\partial f} \langle k^{2} \rangle = \frac{n-1}{n} \frac{\partial F}{\partial f} \Rightarrow \langle k^{2} \rangle = -\langle n-1 \rangle (\phi_{0})^{2} + 2 \langle n-1 \rangle \left[ \frac{\sum a_{ij}^{2} + b}{1 + e^{4J_{ij} + 2b}} + \frac{1}{2} \frac{n}{n-1} \right] \Rightarrow \langle k^{2} \rangle = -\langle n-1 \rangle (\phi_{0})^{2} + 2 \langle n-1 \rangle \left[ \frac{\sum a_{ij}^{2} + b}{1 + e^{4J_{ij} + 2b}} + 1 \right] \Rightarrow \langle k^{2} \rangle = -\langle n-1 \rangle (\phi_{0})^{2} + 2 \langle n-1 \rangle \phi_{0} + \frac{1}{2} \langle n-1 \rangle \frac{1}{f}.
\]

Therefore:

\[
H = -n(\langle n-1 \rangle (\phi_{0})^{2} + 2 \langle n-1 \rangle \phi_{0} + n(\langle n-1 \rangle B_{\phi_{0}}
\]

\[
= n(\langle n-1 \rangle \phi_{0} + [J_{ij} + 2J + B] = n(n-1) \phi_{0} + [J_{ij} + 2J \phi_{0} + B = \tanh^{-1}(2\phi_{0} - 1) - 2J \phi_{0}]
\]

\[
= \sum_{i=1}^{n} \left[ 2J - 3J \phi_{0} + \tanh^{-1}(2\phi_{0} - 1) \right] \frac{1}{2} \sum_{i=1}^{n} k_{i}
\]

\[
= 2 \left[ 2J - 3J \phi_{0} + \tanh^{-1}(2\phi_{0} - 1) \right] \frac{1}{2} \sum_{i=1}^{n} k_{i}
\]

\[
= 2 \left[ 2J - 3J \phi_{0} + \tanh^{-1}(2\phi_{0} - 1) \right] \frac{1}{2} \sum_{i=1}^{n} k_{i}.
\]

So, we end up:

\[
H = \Theta m(G), \quad \Theta = \ln\phi_{0} + 2(2 - 3\phi_{0})J. \tag{23}
\]

This is a significant conclusion for two reasons. Firstly, we were able to find an approximate solution in the 2-star model. Secondly, the specific solution found can be easily integrated into the factorization process, thus displaying the entities that create the spam [14, 17].

Regarding the probability that models the local properties of the graph edges, in the probabilistic NMF we propose, the Poisson distribution optimizes the generalized Kullback–Leibler deviation. Thus, the following function greatly simplifies the required calculations [24, 44, 47]:

\[
p(a_{ij}[a_{ij}] \alpha_{ij}^{a_{ij}} e^{-a_{ij}}. \tag{24}
\]

Having chosen the probability distribution and the ex-probability, we can now use the classical relation of the Bayesian inference to approach the ex post probability that the parameters of our model (the elements of the table \( A e = WH \) take specific values, as well as the hyper-parameters \( \Theta \) of the model [48].

For the proposed model to succeed in imitating the actions of the human brain in a simplified way, the attention mechanism is used, which is also an attempt to implement
the same measure of selective human concentration in some relevant things while ignoring some, respectively. This procedure allows for the particular treatment of different versions of the same situation and identifying events that significantly change the proposed control process. An abstract implementation related to the environment vector \( c_i \) for output \( y_i \) is generated using the weighted sum of the annotations so that [44]:

\[
c_i = \sum_{j=1}^{T} a_{ij} h_j.
\] (25)

In the simplest case, the weights \( a_{ij} \) are calculated from a softmax function given by the following equation:

\[
a_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{r} \exp(e_{ik})},
\] (26)

where \( e_{ij} \) is the output rating of the feed mechanism described by function \( a \), which attempts to record the alignment between the input to \( j \) and output to \( i \).

So given that the element \( a_{ij} \) results from the interior product of the \( i \)th row vector of the table \( W \) on the \( j \)th column vector of \( H \), we arrive at the following relation [40, 44, 49]:

\[
P(w_i^T h_j | a_{ij}, \Theta) \propto \frac{w_i^T a_{ij}}{a_{ij}}! e^{(\alpha-1)w_i^T h_j}.
\] (27)

The slope of the equation for the vectors \( w_i^T h_j \) is calculated as follows:

\[
\nabla D_{w_i}(a_{ij}, w_i^T, h_j) = \sum_{j=1}^{k} \left[ -\frac{a_{ij}}{w_i^T h_j} w_i^T h_j + (1 - \Theta) h_j^T \right]
\]

\[
= \frac{a_i^T}{a_1} h_j^T + (1 - \Theta) e^T H^T,
\] (28)

\[
\nabla D_{h_j}(a_{ij}, W_i^T, h_j) = \sum_{j=1}^{k} \left[ -W_i^T + \frac{a_{ij}}{W_i^T h_j} + (1 - \Theta) W_i^T \right]
\]

\[
= -W_i^T a_i^T + (1 - \Theta) W_i^T e^T,
\]

while, respectively, they are renewed as follows:

\[
(W_i^T)^{(t+1)} \leftarrow (W_i^T)^{(t)} \frac{\nabla D_{W_i}^T}{\nabla D_{W_i}} \rightarrow (W_i^T)^{(t+1)} \leftarrow \frac{a_i}{a_1} H_i^T H_i^T + (1 - \Theta) e^T H_i^T
\]

\[
h_j^{(t+1)} \leftarrow h_j^{(t)} \frac{\nabla D_{h_j}^T}{\nabla D_{h_j}} \rightarrow h_j^{(t+1)} \leftarrow \frac{W_i a_i}{(1 - \Theta) W_i^T e^T}
\] (29)

resulting in the following information rules for the base tables and coefficients \( W \) and \( H \):

\[
W^{(t+1)} \leftarrow W^{(t)} \frac{A/WH H^T}{(1 - \Theta) E^T H^T},
\]

\[
H^{(t+1)} \leftarrow H^{(t)} \frac{W^T A/WH}{(1 - \Theta) W^T E}
\] (30)

Thus, in combination with the deterministic calculation of \( \Theta \), we have the following:

\[
1 - \Theta > 0 \Rightarrow \Theta < 1 \Rightarrow \ln \phi_0 + 2(2 - 3\phi_0) J < 1.
\] (31)

In conclusion, all social networks are far from being characterized as cliques, so the value of connectivity \( \phi_0 \) is much lower than unity and closer to zero (i.e., \( \phi_0 \ll 1 \) applies). Therefore, the term \( \ln \phi_0 \) takes very small (negative) values. On the other hand, the hyperparameter \( J \) is positive (its use in the Hubbard–Stratonovich transform imposes its nonnegativity), so the second term of the sum is positive (it is an exact mathematical transformation that is used to convert a particle theory into its respective field theory by linearizing the density operator in the many-body interaction term of the Hamiltonian and introducing an auxiliary scalar field [50]). Thus, in social network graphs, the effect of the term \( \ln \phi_0 \) is stronger in shaping the final value of the hyperparameter \( \Theta \), so the above relation applies to all networks where the original hypotheses investigated apply. The contribution of \( \phi_0 \) to factorization (more specifically of ERGMs) and its contribution to the aggregation of members of a user’s network to generate recommendations are to be quantified in the following experimental process.

### 4. Scenarios and Results

To model an email referral system, this study uses user actions about activity and the likelihood that a node will engage in abnormal behavior related to the spread of spam. The application is used based on a case study applied in the environment of educational organizations. Instead of classifying two classes into junk and desirable, we treat the problem as multi-category classification in which each class is a user recommendation action in an email. The most common activities are reading, replying, saving, waiting, deleting from the mailing list, terminating, including junk mail filters, and deleting. As part of a collaborative social filtering system, the previously described algorithm was applied to two different datasets containing user actions in email messages and information about their social relationships.

The first collection is small and consists of usage data and, more specifically, the relevant statistics of a user’s traffic to an educational podcast service. On the other hand, the second collection is medium-sized and contains educational material evaluations on an online learning platform on a five-point scale. Despite their differences, the two datasets...
are highly sparse and show the characteristics of free-scale networks regarding the number of ratings they contain and the distribution of social network nodes (most edges fall on a few nodes, while most nodes touch a few edges). A user’s ratings are extracted from the data collection and divided into two distinct sets: training and testing. The test set is then repositioned in the data collection. In the next step, the algorithms generate recommendations, and a list of objects is returned as output, compared with the data in the test set. The whole process is repeated five times for a list size of 5 to 25 items. For the generated recommendations to make sense, in each iteration of the experimental protocol, only users are selected who have evaluated at least twice the number of objects from the respective list size. Several memorandum collaborative systems were implemented to assess the quality of the recommendations produced in different environments.

The levels of the metrics refer to the average of the respective values, while the similarity function used was the logarithm of the probability ratio of the data. Local-level trust metrics were also examined, namely the MoleTrust 1 algorithm (which considers only the users with whom the user in question is directly associated with acne) and the MoleTrust 2 (which also finds the neighbors of the user in question). TidalTrust was also implemented, which calculates user similarity based on the shortest paths between the user in question and all other users on the same connected component. The comparison also includes an algorithm that calculates each user’s reputation on the social network and, more specifically, TrustWalker, which takes a random walk on the graph, selecting its next step uniformly at random. The random walk starts from the user in question, and when it reaches its stationary distribution, the nodes with the highest probability are returned as more similar. Finally, the user network clustering methodology was tested to estimate the relative performance of ERGMs in the position of ex ante probability. The configuration includes the Bayesian NMF algorithm using the Poisson distribution in place of the likelihood. The method of filtering the neighbors that were applied was that of the nearest-; i.e., in producing the recommendations, only the nearest neighbors are taken into account. The value of parameter was set to 5 after a series of verification experiments in which it was found that for values of less than 5, the results of the metrics were unstable. In contrast, the results were lower for values of greater than 5. It should be noted, however, that the relative classification of the algorithms remains constant, regardless of the value of [20, 23, 31, 44, 51].

A first observation of the results is that the competing algorithms perform significantly lower performance than the proposed one. This behavior is attributed to the fact that the proposed methodology’s interaction between users and objects is deemed valid and can be the basis for a possible recommendation. Another interesting point is that the density ratio between the two datasets is reflected almost linearly in the individual results achieved by the different systems. An equally important observation is that each user’s network is a good source of information for making recommendations in sparse datasets, clearly superior to traditional collaborative methods. As the level of data dilution increases, the efficiency of conventional collaborative recommendation algorithms decreases, as there are fewer and fewer users whose object ratings match. In this case, social algorithms “unfold” all their dynamics. The results prove the apparent superiority of the social algorithms that perform a local search on all metrics. It is also worth noting that all three algorithms based on local social network search show similar results even though they explore each user’s neighborhood at a different depth.

Nevertheless, they do not perform well in either dataset. The above observation concludes that basing the recommendations exclusively on the social network’s most popular (or the most frequently visited) nodes does not guarantee similarity in preferences. Finally, the proposed methodology seems to achieve the best results compared with the local and the full search on the social network. Placing the nodes that are part of a user’s network in overlapping clusters leads to complete analysis of users’ proximity to the network, especially when compared to the basic assumptions made by other algorithms. The addition of ERGMs further improves this analysis in place of the ex ante probability. In this way, the central view given to each acne by the Bayesian NMF is addressed, i.e., by introducing structural features of the graph in the process.

5. Conclusions

The consequent increase in the popularity of online educational resources, combined with this lack of preparedness, has made the education sector an ideal target for digital phishing attacks. The detection of spam and the timely assessment of these threats allow the detection of events that can significantly mitigate the effects of organized cyber-attacks. An advanced deep attention collaborative filter was presented to ensure the educational processes and the protection of the educational system. It is a specialized application of intelligent techniques where the spam problem is examined as a social graph to identify harmful communities. Using a deep attention mechanism, the methodology becomes personalized for each user, while critical, innovative optimization processes help evaluate social graphs.

This methodology is being tested with great success in implementing mail protection systems for educational organizations based on the integration of mnemonic, social, and collaborative recommendation systems. The results obtained show a steady improvement of all performance metrics compared with all comparable implementations. This improvement is attributed to the unique filtering capabilities of the proposed methodology. Instead of summing up many users, it tries to discover patterns in their social behavior by (overlapping) grouping them into regions. This process bears similarities to how overlapping finding commonalities algorithms are operating.

A possible research direction would be to include factorizing more complex network features, such as triangles. In this case, however, the model calculations become pretty complicated. It is not easy to derive an approximate solution
similar to the one presented, so the implementation with advanced equipment such as GPU or TPU should be investigated. Accordingly, a more generalized approach to model estimation should be explored as to whether they would achieve better results and greater generalization in community evaluation. Finally, adopting post-hybrid methods of automatic optimization of communities is considered very important for the further development and use of the methodology.

Data Availability
The data are available on reasonable request.

Conflicts of Interest
The authors declare no conflicts of interest.

References

[1] W. Holmes and S. Anastopoulou, “What do students at distance universities think about AI?” in Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale, pp. 1–4, Chicago IL USA, June. 2019.

[2] N. A. Khan, S. N. Brohi, and N. Zaman, “Ten deadly cyber security threats amid COVID-19 pandemic,” May 2020.

[3] T. Ahmad, “Corona virus (COVID-19) pandemic and work from home: challenges of cybercrimes and cybersecurity,” SSRN Electronic Journal, 2020.

[4] J. Chigada and R. Madzinga, “Cyberattacks and threats during COVID-19: a systematic literature review,” SA Journal of Information Management, vol. 23, no. 1, February. 2021.

[5] T. Weil and S. Murugesan, “IT risk and resilience-cr,” IT Professional, vol. 22, no. 3, pp. 4–10, May 2020.

[6] M. Iezzi, “Practical privacy-preserving data science with homomorphic encryption: an overview,” in Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), pp. 3979–3988, Atlanta, GA, USA, December 2020.

[7] R. Khweiled, M. Jazzar, and D. Eleyan, “Cybercrimes during COVID-19 pandemic and work from home: challenges of cybercrimes and cybersecurity,” SSRN Electronic Journal, 2020.

[8] J. Chigada and R. Madzinga, “Cyberattacks and threats during COVID-19: a systematic literature review,” SA Journal of Information Management, vol. 23, no. 1, February. 2021.

[9] T. Weil and S. Murugesan, “IT risk and resilience-cr,” IT Professional, vol. 22, no. 3, pp. 4–10, May 2020.

[10] M. Iezzi, “Practical privacy-preserving data science with homomorphic encryption: an overview,” in Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), pp. 3979–3988, Atlanta, GA, USA, December 2020.

[11] B. Deng, X. Zhang, W. Gong, and D. Shang, “An overview of extreme learning machine,” in Proceedings of the 2019 4th International Conference on Control, Robotics and Cybernetics, pp. 189–195, CRC, Tokyo, Japan, September. 2019.

[12] S. Raschka, “An overview of general performance metrics of binary classifier systems,” October. 2014, http://arxiv.org/abs/1410.5330.

[13] P. Bevington and D. K. Robinson, Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill Education, New York, 2003.

[14] A. A. Abdullahi and M. Kaya, “A deep learning based method to detect email and SMS spams,” in Proceedings of the 2021 International Conference on Decision Aid Sciences and Application (DASA), pp. 430–435, Sakheer, Bahrain, December. 2021.

[15] Y. Fang, C. Zhang, C. Huang, L. Liu, and Y. Yang, “Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism,” IEEE Access, vol. 7, pp. 56329–56340, 2019.

[16] S. Phomkeona and K. Okamura, “Zero-day malicious email investigation and detection using features with deep-learning approach,” Journal of Information Processing, vol. 28, no. 0, pp. 222–229, 2020.

[17] S. Kaddoura, O. Alfandi, and N. Dahmani, “A spam email detection mechanism for English language text emails using deep learning approach,” in Proceedings of the 2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), Bayonne, France, pp. 193–198, Bayonne, France, September. 2020.

[18] D. S. Asudani, N. K. Nagwani, and P. Singh, “Exploring the effectiveness of word embedding based deep learning model for improving email classification,” Data Technologies and Applications, February. 2022.

[19] N. Ali, A. Fatima, H. Shahzadi, A. Ullah, and K. Polat, “Feature extraction aligned email classification based on imperative sentence selection through deep learning,” Journal of Artificial Intelligence and Systems, vol. 3, no. 1, pp. 93–114, 2021.

[20] C. C. Aggarwal, “Neighborhood-based collaborative filtering,” in Recommender Systems: The Textbook, pp. 29–70, Springer International Publishing. New York, 2016.

[21] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, “Recommendation systems: principles, methods and evaluation,” Egyptian Informatics Journal, vol. 16, no. 3, pp. 261–273, 2015.

[22] M. Deschênes, “Recommender systems to support learners’ Agency in a Learning Context: a systematic review,” International Journal of Educational Technology in Higher Education, vol. 17, no. 1, p. 50, October. 2020.

[23] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, “Collaborative filtering recommender systems,” in The Adaptive Web: Methods and Strategies of Web Personalization, pp. 291–324, Springer, Berlin, Heidelberg, 2007.

[24] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” Nature, vol. 401, no. 6755, pp. 788–791, October. 1999.

[25] N. Thai-Nghe, L. Drumond, A. Krohn-Grimberge, and L. Schmidt-Thieme, “Recommender system for predicting student performance,” Procedia Computer Science, vol. 1, no. 2, pp. 2811–2819, 2010.

[26] M. Srafi, A. Oussous, A. Ait Lahcen, and S. Mouline, “Recommender systems based on collaborative filtering using review texts-A survey,” Information, vol. 11, no. 6, pp. 317–6, June. 2020.

[27] I. A. A–Q. Al-Hadi, N. M. Sharef, N. Sulaiman, and M. Mustapha, “Review of the temporal recommendation system with matrix factorization,” International Journal of Innovative Computing, Information and Control, vol. 13, p. 16, 2017.

[28] S. Gupta, S. Al-Obaidi, and L. Ferrara, “Meta-analysis and machine learning models to optimize the efficiency of self-healing capacity of cementitious material,” Materials, vol. 14, no. 16, p. 4437, January. 2021.

[29] S. Joshi, T. Jain, and N. Nair, “Emotion based music recommendation system using LSTM - CNN architecture,” in Computational Intelligence and Neuroscience
[30] G. Alexandridis, G. Siolias, and A. Stafylopatis, "Accuracy versus novelty and diversity in recommender systems: a nonuniform random walk approach," in Recommendation and Search in Social Networks, O. Ulusoy, A. U. Tansel, and E. Arkun, Eds., Springer International Publishing, New York, NY, pp. 41–57, 2015.

[31] J. O. Berger, "Bayesian analysis," in Statistical Decision Theory and Bayesian Analysis, pp. 118–307, Springer, New York, NY, 1985.

[32] M. S. Mahmud, J. Z. Huang, S. Salloum, T. Z. Emara, and K. Sadatdiyinov, "A survey of data partitioning and sampling methods to support big data analysis," Big Data Mining and Analytics, vol. 3, no. 2, pp. 85–101, 2020.

[33] J. Chanda, S. Sengupta, A. Kanjilal, and S. Bhattacharya, "CA-graph: a context aware graph to model enterprise service bus," in Proceedings of the 2010 Annual IEEE India Conference (INDICON), pp. 1–4, Kolkata, India, December 2010.

[34] M. Gori, M. Maggini, and L. Sarti, "Graph matching using random walks," in Proceedings of the 17th International Conference on Pattern Recognition, vol. 3, pp. 394–397, Cambridge, UK, August 2004.

[35] B. K. Deka, "Transformations of graph database model from multidimensional data model," in Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 2836–2840, New Delhi, India, March. 2016.

[36] B. He, Y. Cui, J. Chen, and P. Xie, "A spatial data mining method for mineral resources potential assessment," in Proceedings of the 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, pp. 96–99, Fuzhou, China, June. 2011.

[37] S.-Y. Kim and Y.-H. Yoo, "A new haptic model using bond graphs," in Proceedings of the 2009 Fourth International Conference on Computer Sciences and Convergence Information Technology, pp. 387–392, Seoul, Korea (South), November 2009.

[38] A. E. Barinov and A. A. Zakharov, "Clustering using a random walk on graph for head pose estimation," in Proceedings of the 2015 International Conference on Mechanical Engineering, Automation and Control Systems (MEACS), pp. 1–5, Tomsk, Russia, December 2015.

[39] V. Nguyen, J. Leeka, O. Bodenreider, and A. Sheth, "A formal graph model for RDF and its implementation," June. 2016, http://arxiv.org/abs/1606.00480.

[40] S. Blyumin, A. Pogodaev, and E. Khabibullina, "Graph-structural modeling of some special organizational systems," in Proceedings of the 2020 2nd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA), pp. 279–283, Lipetsk, Russia, November 2020.

[41] G. Alexandridis, G. Siolias, and A. Stafylopatis, "An efficient collaborative recommender system based on k-separability," in Artificial Neural Networks – ICANN 2010, pp. 198–207, Berlin, Heidelberg, 2010.

[42] P. MohanaPriya and S. M. Shalinie, "Restricted Boltzmann machine based detection system for DDoS attack in software defined networks," in Proceedings of the 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN), pp. 1–6, Chennai, India, March. 2017.

[43] S. Goel and V. K. Panchal, "Performance evaluation of a new modified firefly algorithm," in Proceedings of the INFOCOM.