Received: 12 August 2018 | Revised: 2 February 2019 | Accepted: 8 March 2019
DOI: 10.1111/exsy.12405

ORIGINAL ARTICLE

WILEY Expert Systems

# Automatic assignment of reviewers in an online peer assessment task based on social interactions

Antonio R. Anaya | Manuel Luque | Emilio Letón | Félix Hernández-del-Olmo

Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain

**Correspondence**
Antonio R. Anaya, Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain.
Email: arodriguez@dia.uned.es

## Abstract

Online peer assessment tasks are very popular and have unique characteristics that improve learning and encourage social interactions in a distance education environment. Unfortunately, social factors have usually been ignored in the process of selecting reviewers for online peer assessment tasks. We hypothesise that this fact could have some influence on the lack of engagement and participation by some learners. For this reason, we propose an approach in which social network analysis techniques, expert criteria, and Bayesian reasoning are applied to select reviewers with the objective of increasing participation in peer review tasks. The approach is divided into two elements. On the one hand, we have developed an influence diagram template that structures a set of proposed social network analysis variables in accordance with expert criteria. This influence diagram template can be easily updated for any course simply by eliciting a minimal set of parameters. On the other hand, we have instantiated the proposed influence diagram template to produce an influence diagram network to quantify the quality of reviewer assignment for an online peer assessment task. In an online experiment, we verified that the consideration of social factors can increase participation in a peer assessment task.

KEYWORDS

e-learning, influence diagram, network analysis, social interactions

## 1 | INTRODUCTION

Online peer assessment (OPA) is the process whereby a learner assesses the work of a coursemate in an online learning environment. OPA is very advisable in online courses because, as Uto and Ueno (2016) note, it helps distance education learners to develop several advantageous capabilities, such as responsibility, autonomy, and self-regulation. Nevertheless, OPA tasks encounter some difficulties (Hanrahan & Isaacs, 2001) that can lead to a lack of engagement and participation by learners (Meek, Blakemore, & Marks, 2016). The characteristics of an OPA can include cooperative learning (Lu, Warren, Jermaine, Chaudhuri, & Rixner, 2015), in which the social interactions between participants play an important role (Bouguessa & Romdhane, 2015; D.Johnson & Johnson, 2005). However, the social factors of peer assessments (e.g., participation in an OPA) still have not been addressed in depth (van Gennip, Segers, & Tillema, 2009). In online course platforms that support OPA tasks, candidates are randomly selected as reviewers and asked to review other coursemates' work (Kao, 2013). These platforms thus ignore the social factors of the course during the reviewer assignment process.

The framework of our research is activity theory, which comprises learners, the activities performed by the learners, the tools they use in these activities, the social relationships that emerge while they are performing these activities, and the outcomes they achieve during these activities (see, e.g, Jonassen, Peck, & Wilson, 1999, Engestrom, 2000, & Deng & Tavares, 2013, among others). Within this framework, the assignment method is a key aspect in encouraging learners to participate in an OPA task. According to Deng and Tavares (2013), the established social relationships encourage learners to participate. Our hypothesis is that considering social factors in the selection of reviewers will improve participation in an OPA task.

We propose a reviewer assignment method that considers the social factors of a common e-learning experience. The objective is to encourage participation in the OPA task. First, we designed a template to specify the structure and parameters of our approach. Second, we instantiated the proposed template to generate a procedure using learners' interaction data from the online Bachelor's Degree in Computer Engineering subject

  

"Complexity and Computability" offered by the distance education university UNED. Third, we put an OPA task for this subject into practice to test our hypothesis, which is defined as follows: considering social factors during the reviewer assignment process will increase participation in OPA tasks. The details of our study are as follows:

1. Template design:

    (a) We identified social variables. Social network analysis (SNA; Scott,2017), which is the process of capturing and analysing social relations through the use of graph theory, is an emerging research area that has been applied in several fields (Aggarwal, 2011; Wasserman & Faust, 1994). In our research, we used SNA to obtain social information such as learner popularity, activity, and proximity for consideration when assigning reviewers.
    (b) We structured and ordered the social variables. We took the experience of distance education teachers into consideration by using such teachers as experts from whom we elicited knowledge for our approach. We asked them for their preferences with regard to the assignment of learners for peer assessment. Thus, the experts provided us with a set of criteria that could be used to order and weight the SNA variables.
    (c) We elicited the parameters of the template. In an e-learning environment, there are various hidden or uncertain variables that can influence learners' behaviour (Anaya, Luque, & García-Saiz, 2013). A Bayesian network (BN; Pearl1988) is a type of framework for performing reasoning on problems that are subject to uncertain and probabilistic factors. Influence diagrams (IDs; Howard & Matheson1984) extend BNs to enable automatic decision-making under uncertainty. Therefore, we built an ID template in which the identified SNA variables are structured in accordance with the collected expert criteria but independently of any specific course. Because this template is independent of the set of learners and the course subject, it can be easily updated for application to other learning contexts.

2. Implementation of a real OPA task:

    (a) We obtained the data used to account for social factors from forum interactions related to the online course for the above-mentioned subject.
    (b) We performed SNA with the collected interaction data.
    (c) With the SNA results, we instantiated the ID template to produce an ID network to quantify the quality of the assignment of the reviewers.
    (d) We implemented a computational process to finds a specific reviewer assignment for each learner based on the ID results.

3. Experimentation:

    (a) Finally, we compared our reviewer assignment method with a random assignment method, which is commonly used in OPA tasks (Kao, 2013). Our objective in the experiment was to test whether our reviewer assignment method encouraged students to participate in the OPA task. We partitioned the set of learners into two groups. In the *experimental group*, reviewers were assigned using our proposed method. In the *control group*, reviewers were assigned randomly. We measured learner participation and achieved appreciable improvement in learner participation in the first group.

Following our hypothesis, we first introduced social factors into the formulation of the reviewer assignment problem using a set of SNA variables (centrality metrics) that are well known in the research literature (Dawson, 2008). Second, we introduced weighting based on expert experience to structure the SNA variables into an ID template. The designed template is independent of the specific characteristics of the e-learning course and thus can be easily adapted to different e-learning contexts. We showed that our proposed assignment method more strongly encourages learners to participate in an OPA task than a random assignment method does.

In the following section, we provide a general description of the advantages of using OPA tasks and review some research on improving them. In Section 3, we define in depth the purpose of our designed template for assigning reviewers while considering social factors. Next, we describe the procedure for using the proposed template design in the implementation of a real OPA task. To prove our hypothesis of improved participation in an OPA task when social factors are considered, we describe the experiment that we performed in Section 5. Finally, we briefly discuss our conclusions and plans for future work.

## 2 | THE CONTEXT OF OPA TASKS

### 2.1 | Peer assessment in online education

The main advantages of peer assessment have been well known since the end of the last century: motivation for learners and the fact that feedback from other learners is readily understood. Consequently, learning is more effective with peer assessment (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Falchikov, 1986; K.Topping, 1998, among others). With the help of modern technologies in the education environment, peer assessment has been extensively applied, mainly in online education (K. J. Topping, 2005). OPA also has other advantages (see, e.g., Gielen & De

Wever, 2015;Rosa, Coutinho, & Flores, 2016; Ueno & Okamoto, 2008; Uto & Ueno, 2016) in reducing teacher workload in both the assessment and feedback processes (Bouzidi & Jaillet, 2009) and increasing studentŰstudent interaction (Rosa et al., 2016). Because of these advantages, OPA has been incorporated into new types of online courses, such as massive open online courses (MOOCs), small private online courses, small open online courses, corporate open online courses, and nano open online courses.

To achieve a proper assessment in OPA, it is crucial for the learner to have acquired thorough knowledge of the subjects covered (Ballantyne, Hughes, & Mylonas, 2002). It is known that a key factor in achieving this goal is learners' online participation (see, e.g., ; Hoskins & Van Hooff, 2005, Huang, Lin, & Huang, 2012, and among others). This online participation can be influenced by several factors (Deng & Tavares, 2013):

- *Individual*: It is difficult for a learner to participate if he/she is not familiar with the subject under discussion.
- *Social*: Established social relationships encourage learners to participate.
- *Pedagogical*: If learners perceive the discussion to be interesting, then they are motivated to take part in it.
- *Technological*: The ease of use of the platform influences learners' decisions to participate in online activities.

At present, however, learners are usually less excited to participate on the forums of a learning management system (LMS) than on social networks (Deng & Tavares, 2013) because it is difficult to participate when there are no previously established relationships (Cheung, Hew, & Ng, 2008). Bearing these factors in mind, we performed some activities on Twitter to foster the social environment of and promote participation in an LMS (Deng & Tavares, 2013) to prepare learners for an appropriate OPA task.

Another crucial aspect for the designers of an OPA to consider is that they need to ensure that learners perfectly understand all criteria. They also need to transmit to the learners a sense of membership in their peer learning to ensure a fair assessment (see, e.g., Hovardas, Tsivitanidou, & Zacharia, 2014).

### 2.2 | Research on peer assessment

Above, we have seen the advantages of OPA, but for some learners, an OPA task could be difficult and time-consuming (Hanrahan & Isaacs, 2001). This difficulty can lead to a lack of effort (Lu et al., 2015), engagement, and participation (Meek et al., 2016). Some research has addressed this issue with the aim of improving engagement and participation in OPA tasks.

Recently, OPA has become a common and standard mode of assessment in MOOCs mainly due to the massive number of students involved in an e-learning context. However, is it appropriate or suitable for these contexts? Meek et al. (2016) studied this issue by examining learner participation, performance and opinions in the context of OPA tasks. The objective of their research was to study under what conditions an OPA task is suitable in an online education environment. They found that learners' opinions on the usefulness of the OPA task were mixed: some strongly believed it benefitted their learning, whereas others did not find it useful or did not participate. They suggested instructional design strategies to improve student participation and increase learning gain.

Moreover, Lu et al. (2015) conducted a controlled study to examine how learners might be motivated to do a better job during an OPA task. Their goal was to identify a simple and practical method of motivating students to perform high-quality peer assessment and then to rigorously test the identified method in a controlled experiment. They proposed a method of evaluating reviewer assessments in OPA tasks and informing the reviewers of their evaluations. These authors proposed imposing personal accountability on the reviewers in accordance with the basic element of cooperation identified by D. W. Johnson and Johnson (2004).

Personal accountability for each learner and relations of positive interdependence between learners were integrated into the approach proposed by Kao (2013), which encouraged learners to adopt a broader and more comprehensive perspective during peer assessment, enhancing their ability as "critical assessors." Kao (2013) also noted that most online learning environments that support OPA tasks use a process in which learners are randomly assigned as reviewers.

Very few online learning environments have applied assignment criteria different from random assignment. Capuano and Caballé (2015) proposed a smart assignment procedure based on learners' past grades in previous peer assessment tasks. The authors focused on the learners' certainty in the OPA task as a main factor in the assignment process. These authors measured a learner's certainty as the similarity between the assessment given by the learner and the assessment given by the teacher. They proposed a method that used this certainty to weight the learners for the assignment process.

## 3 | DESIGN OF THE SOCIAL ANALYSIS-BASED REVIEWER ASSIGNMENT METHOD

We have designed a reviewer assignment method that considers social factors in the learning process. In our educational context, the LMS forums are the main means of learner socialisation. We extracted SNA variables characterising the learners' social behaviour from their interactions on the forums. To identify criteria for selecting the most appropriate learner for a peer assessment, we asked experts to guide the decision process of our reviewer assignment method. We then proposed an ID template (Howard & Matheson, 1984) to consider these expert criteria for assigning learners for an OPA task. Figure 1 shows the corresponding schema. SNA was applied to propose a set of well-known variables. The expert criteria indicated a method these variables in terms of their importance. We built the ID template by combining the variables with the expert criteria.

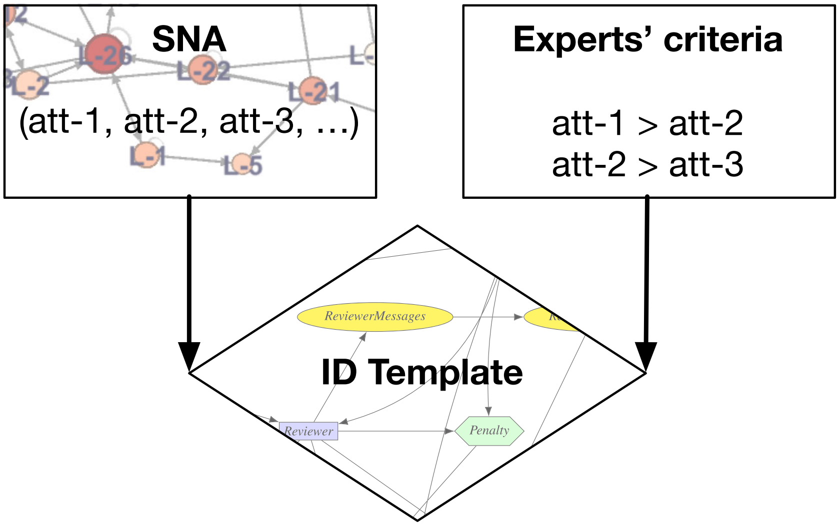


**FIGURE 1** Reviewer assignment method design schema

## 3.1 | Method variables

In the context of online education, it is common to have an LMS that supports learners by providing services that can be used to communicate or interact. The most important and popular of these services are the forums. Several researchers have analysed forums to obtain indicators of student actions (Daradoumis, Juan, Lera-López, & Faulin, 2010). On the one hand, we extracted the learners' interactions on the forums to characterise the learners' social behaviour. We focused on simple variables that could explain learner behaviour in a communication network, and we followed a simple statistical approach to infer two learner features (Anaya & Boticario, 2011):

**The number of forum messages exchanged by two learners** This variable captures the real interactions between two learners. We call this variable SharedMessages.

**The number of conversations that a learner has started on the forums** A high value in this attribute indicates that a learner often initiates new conversations on the forums. We call this variable Initiative.

On the other hand, we applied SNA to the extracted forum interactions to infer the social relations among users (Dawson, 2008). SNA techniques are commonly used to infer learners' behaviours in relation to others (Sinha, 2014). For this reason, we analysed the forum interactions to discover the learners' popularity, relevancy, closeness, etc. in the group. We considered the learners as nodes of a directed graph and the interactions between two learners as an arc connecting the two corresponding nodes. We used the following SNA metrics associated with each node:

**Outdegree** In our context, this metric measures the number of messages that a learner has sent. From this attribute, we obtain the variable Messages. However, this attribute is only roughly related to the activity of the learner (Zafarani, Abbasi, & Liu, 2014). Although we have recorded the variable messages, it is reasonable to think that the number of messages is not the only factor that influences the level of activity of a student. Other indicators could also provide information about the level of activity of the learner. Thus, there is an uncertain relationship between the number of messages and the activity of the learner. We propose to solve this issue by using a probabilistic relation in the ID, as we explain in depth in Section 3.3.2. Thus, from this attribute, we obtain two variables describing a learner: Messages and Activity.

**Indegree** In our context, this metric measures the number of replies to messages sent by a learner. According to Zafarani et al. (2014), this attribute indicates the popularity or reputation of the learner. We call this variable the Popularity of the learner.

**Closeness centrality** Intuitively, the nearer a learner is to other learners (the higher his or her closeness centrality is), the faster he or she will accede to other learners (Zafarani et al., 2014). We call this variable the Closeness of the learner.

**Betweenness centrality** Intuitively, when a learner more frequently acts as an infomediary (a person who disseminates information among neighbours) between two other learners that learner will have a higher betweenness (Zafarani et al., 2014). We call this attribute the Betweenness of the learner.

**Eigenvector centrality** The eigenvector centrality is a relative score assigned to each node in the network based on the concept that connections to high-scoring nodes contribute more than others. Therefore, the more important or connected the neighbours of a learner are, the higher that learner's eigenvector centrality value will be (Zafarani et al., 2014). We call this attribute the Eigenvector of the learner.

The seven variables proposed here are easy to calculate, easy to understand by a neophyte, and easy to apply in other contexts.

### 3.2 | Expert criteria

Once we had selected the set of variables to be used to characterise the learners' social behaviours, we needed a set of criteria for evaluating their importance. We used teachers as experts and elicited knowledge from them. We interviewed the teachers of the online learning course to which we applied our method. The interview was structured around the selected SNA variables and the learner behaviour that these variables can reveal (Sagheb-Tehrani, 2006). The teachers explained to us their experiences in assigning learners for OPA tasks, and they proposed the following set of criteria:

1. It is advisable for the activity levels of learners participating in the OPA task to be compatible. We can approximate learner activity by means of the variable Messages. We defined the variable ActivityCompatibility to account for this compatibility. In the next section, we explain this variable in depth.
2. It is advisable for the reviewer and the learner he or she is assessing to know each other. We can determine whether two learners have contacted each other by means of the variable SharedMessages, which is a numeric variable with a value equal to the number of messages exchanged between the two learners.
3. The learners' social variables should be arranged in order of increasing importance as follows: Popularity, Eigenvector, Betweenness, Initiative, and Closeness. This list expresses that learner Closeness is more important than learner Initiative, learner Initiative is more important than learner Betweenness, and so on.

We asked the experts to sort all criteria according to their importance. In addition to the order of importance of the social characteristics of the learners, as described above, the experts told us that ActivityCompatibility is more important than SharedMessages, which, in turn, is more important than any other social learner variable. Thus, the criteria can be listed in order of increasing importance as follows: Popularity, Eigenvector, Betweenness, Initiative, Closeness, SharedMessages, and ActivityCompatibility. Therefore, the least important criterion is Popularity, and the most important criterion is ActivityCompatibility.

### 3.3 | Influence diagram

We are interested in finding an approach for selecting students for an OPA task that can consider social factors. Such an approach should provide us with a method of automatically deciding which students are most appropriate to act as reviewers of a particular student. For this purpose, we recorded several variables based on SNA and forum interactions (see Section 3.1). Additionally, the expert criteria described above gave us a rough idea of the relative importance of these variables in selecting reviewers. We then needed to implement a computational system to automatically accomplish the goals of our approach.

One consideration is that to solve the problem, we must deal with uncertainty. For example, we recorded the variable StudentMessages; however, it is reasonable to think that the number of messages is not the only factor that influences the level of activity of a student. Other indicators could also provide relevant information in this regard. However, these other indicators are outside of the control of the decision maker. Accordingly, we need a framework in which this lack of certainty can be accounted for.

IDs (Howard & Matheson, 1984) constitute a graphical framework for extending BNs (Pearl, 1988) to enable automatic decision making under uncertainty. IDs compactly represent the probabilistic and functional dependences among the variables of a decision problem. Therefore, we built an ID for reviewer selection for an OPA task.

Formally, the structural part of an ID consists of an acyclic-directed graph defined over three types of variables: *chance* variables, which represent uncertain events that are outside of the control of the decision maker; *decision* variables, which represent actions controlled by the decision maker; and *utility* variables, which quantify the preferences of the decision maker. In the context of IDs, the terms variable and node (of the graph) are typically used interchangeably.

We distinguish two types of utility nodes: (a) *ordinary* utility nodes, whose parents are chance or decision nodes and (b) *super-value* utility nodes (Luque & Díez, 2010; Tatman & Shachter, 1990), whose parents are other utility nodes.

In an ID, arcs into chance nodes represent probabilistic dependence. Arcs into decision nodes represent the availability of information, that is, an arc from a variable $X$ to a decision node $D$ indicates that the value of $X$ is known when deciding on $D$. Arcs into utility nodes represent functional dependence.

The quantitative part of the ID consists of assigning a function to each chance or utility node: (a) a probability function for each chance variable, (b) a real-valued function for each ordinary utility node, and (c) an additive or multiplicative combination of functions for each super-value utility node.

#### 3.3.1 | Structure of the graph

The graph of the ID (see Figure 2) was built on the basis of the collected expert knowledge of the cause–effect and probabilistic relations of dependence and independence among the variables. Below, we list all nodes in the diagram and group them into the three types of ID nodes: chance, utility, and decision.

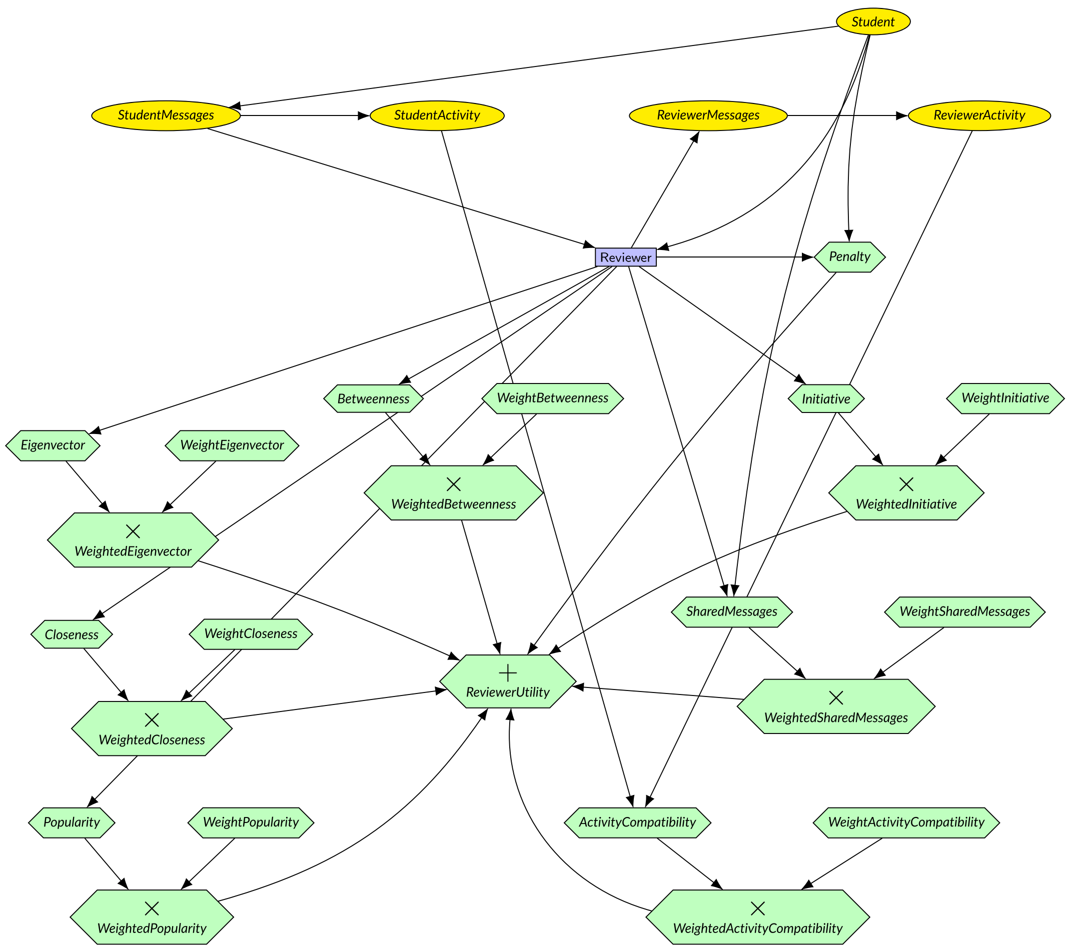


**FIGURE 2** Influence diagram model

### Chance nodes

- Student: This node represents the student who is being considered at a given moment. It is discrete and has a state corresponding to each student in the course.
- StudentMessages: This node corresponds to the variable Messages (explained in Section 3.1) for each student. The variable StudentMessages has a deterministic relationship with Student, as there is no uncertainty in the relationship between these two variables. Thus, if the value of the variable Student is known with certainty, then the value of the variable StudentMessages is also known with certainty. The domain of this variable contains three states: low, medium, and high.
- ReviewerMessages: This node corresponds to the variable Messages (explained in Section 3.1) for each reviewer. The variable ReviewerMessages has a deterministic relationship with the decision variable Reviewer, which will be explained below. The variable ReviewerMessages has the same set of states as that of the variable StudentMessages: low, medium, and high.
- StudentActivity: This node represents the level of activity of each student. Thus, a high value of this variable indicates that the student is very active. The variable StudentMessages is the parent of StudentActivity in the graph; these two variables have a probabilistic relationship because, as we explained at the beginning of this section, although the number of messages is an important factor that influences the level of activity of a student, it is not the only factor; therefore, the level of activity of a student cannot be known with certainty if only the number of messages sent by that student is known. The variable StudentActivity has the same set of states as that of the variable StudentMessages: low, medium, and high.
- ReviewerActivity: This node corresponds to the level of activity of each reviewer. For the same reasons explained above for the variable StudentActivity, the variable ReviewerMessages is the parent of ReviewerActivity in the graph. The variable ReviewerActivity has the same set of states as that of the variable StudentActivity: low, medium, and high.

### Utility nodes

- ActivityCompatibility: This node represents the compatibility in activity level between the student and the reviewer (see Section 3.2). Thus, its parents in the graph are StudentActivity and ReviewerActivity.

- A utility node for each expert criterion presented in Sections 3.1 and 3.2: The graph contains a node for each variable of the statistical approach defined in Section 3.1, namely, Initiative and SharedMessages; a node for each SNA metric variable defined in Section 3.1, namely, Popularity, Closeness, Betweenness, and Eigenvector; and a node for the variable ActivityCompatibility defined in Section 3.2. The utility values of these nodes are real numbers ranging from 0 to 1.
- Nodes with the prefix *Weight* in Figure 2: Each utility node above is associated with a weight utility node that represents a measure of the importance of the corresponding attribute. If we use **A** to denote the set of utility nodes listed above, we can formulate the utility value of a reviewer $r$, denoted by $U(r)$, as a weighted linear combination of the values of the utility nodes in **A**:

$$U(r) = \sum_{X \in \mathbf{A}} U_{weight(X)} \cdot U_X(r), \tag{1}$$

  where $weight(X)$ denotes the utility node corresponding to the weight of attribute $X$. The products and the summation in Equation 1 are also represented by nodes, as we explain below.
- Nodes with the prefix *Weighted* in Figure 2: Each of the utility nodes corresponding to the expert criteria or to ActivityCompatibility is associated with a product node that represents the product of the attribute with its corresponding weight, as expressed in Equation 1. For example, WeightedPopularity is the product of Popularity and WeightPopularity; because of this functional relation, Popularity and WeightPopularity are the parents of WeightedPopularity in the graph, thus making use of the ID extension of super-value nodes proposed by Tatman and Shachter (1990).
- ReviewerUtility: This node is a super-value node that accumulates the sum of the products of each attribute and its weight. Thus, this node represents the summation in Equation 1. This is the only utility node that is a terminal node (a node without children) in the diagram, and it therefore accumulates the global utility function that will be maximised by the inference algorithm used to solve the ID.
- Penalty: This node is a dummy node that was added to the ID to prevent the inference algorithm from assigning a student to him/herself as a reviewer. This node is an addend in the utility function of node ReviewerUtility and takes values from the set $\{0, -\infty\}$. Given two students $x$ and $y$, the utility value of Penalty is 0 if $x \neq y$, and $-\infty$ otherwise.

All ordinary utility variables are numeric.

#### Decision node

- Reviewer: This node represents the reviewer selection decision for each student. Its parents in the graph are the variables whose values are observed when choosing the reviewer, namely, Student, and StudentMessages. The variable Reviewer is discrete and has a state corresponding to each student in the course.

### 3.3.2 | Elicitation of the numerical parameters of the model for a course-independent ID

In this section, we have described the construction of the ID graph, the qualitative part of the ID. Now, to complete the ID, we need to elicit the numerical values corresponding to the probability and utility parameters. However, instead of eliciting all these parameters in a single phase, we can first build an *ID template* by partitioning the set of numerical parameters into two sets:

1. Those parameters that are taken from expert knowledge and thus do not depend on the course to which the model is applied.
2. Those parameters that depend on the course to which the model is applied because they are calculated via the SNA of particular students in that course.

We have built an *ID template* by eliciting only the parameters in Set 1. The main advantage of building this template is that it can be easily updated for application to any other course by eliciting only the parameters in Set 2. The use of this template also improves the quality of the final constructed model because the elicitation of ID parameters is an error-prone task that requires continuous debugging of the inference results returned by the ID (Lacave, Luque, & Díez, 2007).

In the following, we will describe how we elicited the parameters of the ID template. In Section 4.3, we will describe how to instantiate the template by incorporating the parameters in Set 2 for the course to which the model is to be applied, thereby obtaining a complete ID, called an *ID network*.

We wish to emphasise the distinction in our approach between the ID template and the ID network. The use of an ID template has several advantages, as stated above. However, an ID template does not constitute an ID because it is incomplete, as some numerical parameters have yet to be elicited. When the ID template is instantiated with parameters specific to the course to which it is to be applied, that instantiation results in an ID network. This ID network can then be processed with inference algorithms to produce the desired results for our analyses.

#### Elicitation of probability parameters

The probability table of the variable Student is irrelevant as long as it assigns a non-zero probability to each student. Thus, we can simply assign a discrete uniform distribution to this variable.

The probability table of StudentActivity is conditioned on the variable StudentMessages. These two variables are both discrete, and their sets of states are identical. Both variables represent a gradation scale, with their states being ordered in the following sequence: high, medium, and low.

We want the probability table of the variable StudentActivity to satisfy two constraints: if the observed state of StudentMessages is $s$, then (a) $s$ is the most likely state of StudentActivity and (b) the closer another state $z$ is to $s$ on the gradation scale, the higher the probability of Activity = $z$ will be.

Because the variables StudentActivity and StudentMessages are ternary in nature, nine parameters must be elicited to construct the probability table P(StudentActivity | StudentMessages). However, considering the constraints described above, only three of these parameters are independent. We assigned the values listed in Table 1 to the parameters of the probability table. The numbers in bold correspond to the independent parameters.

The probability table of ReviewerActivity is identical to that of StudentActivity, except that the parent of ReviewerActivity in the graph is ReviewerMessages instead of StudentMessages.

### Elicitation of utility parameters

We want the utility table of ActivityCompatibility to satisfy two constraints: if the state of StudentActivity is $s$, then (a) $s$ is the state of ReviewerActivity with the highest utility and (b) the closer another state $z$ of ReviewerActivity is to $s$ on the gradation scale, the higher the utility will be.

Because the variables StudentActivity and ReviewerActivity are ternary in nature, nine parameters must be elicited to construct the utility table of ActivityCompatibility. However, considering the constraints described above, only six of these parameters are independent. Table 2 presents the utility values we elicited for the node ActivityCompatibility.

To elicit the weights of the expert criteria in the utility function, that is, the utility values of the nodes with the prefix *Weight* in Figure 2, we used the overall order of the criteria as provided by the experts. We selected a monotonically increasing function $f : \mathbb{N} \rightarrow \mathbb{R}^{+}$, and then, for each attribute $X$, given its position $i_X$ in the ordered list, we assigned it a weight proportional to $f(i_X)$. The reason for choosing a monotonically increasing function was to guarantee that if an attribute $X$ is more important than an attribute $Y$, then the weight assigned to $X$ will be greater than that assigned to $Y$. We tested three different functions, $f(x) = x$, $f(x) = x^2$, and $f(x) = \sqrt{x}$, whose corresponding weights appear in Table 3.

**TABLE 1** Parameters of P(StudentActivity | StudentMessages)

| StudentMessages | StudentActivity | Probability |
|---|---|---|
| Low | Low | 0.7 |
| Low | Medium | **0.2** |
| Low | High | 0.1 |
| Medium | Low | 0.15 |
| Medium | Medium | **0.7** |
| Medium | High | 0.15 |
| High | Low | 0.1 |
| High | Medium | **0.2** |
| High | High | 0.7 |

*Note*. The numbers in bold correspond to independent parameters.

**TABLE 2** Utility table of ActivityCompatibility

| StudentActivity | ReviewerActivity | Utility |
|---|---|---|
| Low | Low | **0.6** |
| Low | Medium | **0.25** |
| Low | High | **0.15** |
| Medium | Low | **0.2** |
| Medium | Medium | 0.6 |
| Medium | high | 0.2 |
| High | Low | **0.15** |
| High | Medium | **0.25** |
| High | High | 0.6 |

*Note*. The numbers in bold correspond to independent parameters.

**TABLE 3** Weights of the expert criteria in the utility function

| Expert criterion | $f(x) = x$ | $f(x) = x^2$ | $f(x) = \sqrt{x}$ |
|---|---|---|---|
| Popularity | 0.0357142857 | 0.007142857142857143 | 0.0741973329 |
| Eigenvector | 0.0714285714 | 0.02857142857142857 | 0.1049308744 |
| Betweenness | 0.1071428571 | 0.06428571428571428I | 0.1285135503 |
| Initiative | 0.1428571429 | 0.11428571428571428 | 0.1483946657 |
| Closeness | 0.1785714286 | 0.17857142857142858 | 0.1659102801 |
| SharedMessages | 0.2142857143 | 0.257142857142857 | 0.1817456058 |
| ActivityCompatibility | 0.25 | 0.35 | 0.1963076907 |

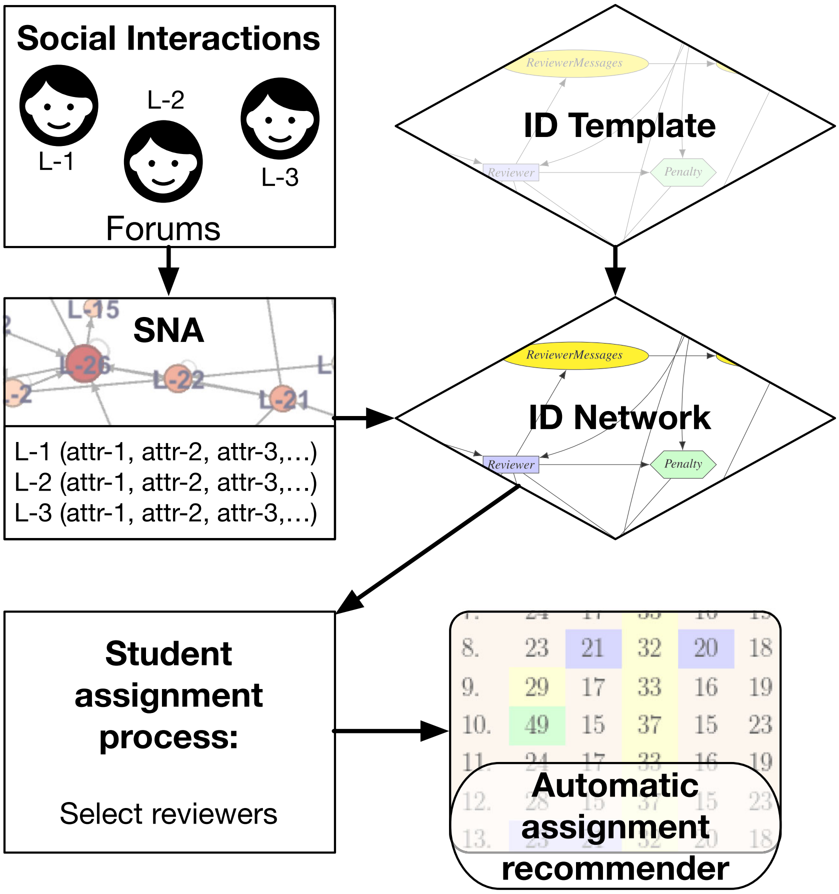


**FIGURE 3** Reviewer assignment method procedure schema

## 4 | PROCEDURE OF THE SOCIAL ANALYSIS-BASED REVIEWER ASSIGNMENT METHOD

We implemented the proposed reviewer assignment method design for use in a real education environment. The procedure that we followed is summarised in Figure 3. We observed forum interactions to extract representative information concerning the social interactions of the learners. With SNA, we deduced the learners' social behaviours. We then instantiated the ID template with the data obtained from the SNA to build an ID network as a quantitative representation of the learners as reviewers. Finally, we used the ID network to assign candidates by maximising the global utility over the set of learners. In the following subsections, we explain in depth the procedure of our reviewer assignment method.

### 4.1 | Educational context

The 2015Ű16 session of the course on the subject “Complexity and Computability”, which is taught in the fourth year of the Bachelor's Degree in Computer Engineering within the framework of the European Higher Education Area, was chosen as the basis for ID instantiation. The course contents focus on the following four topics: introduction to Turing machines, undecidable problems, intractable problems, and other types of problems. The course is based on the book by Hopcroft, Motwani, and Ullman (2007), and e-learning is offered through an LMS provided by UNED, which is called active learning framework .

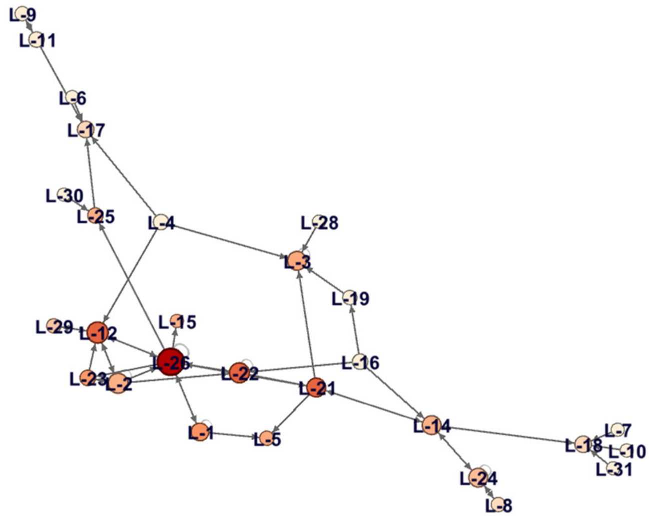


**FIGURE 4** Social learner network-eigenvector centrality

Communication among learners and between learners and teachers occurs via the LMS, through which the *continuous monitoring and evaluation of learning* are conducted. On the one hand, forum assessments are issued, with values from 0 to 10 points depending on relevant participation. On the other hand, students are asked to record a special type of video podcast called a modular teaching mini-video (MTM; see Letón & Molanes-López, 2014 for details) in which they solve true/false questions with reasoned responses by giving proper demonstrations or counterexamples. MTMs have a feature of animation that has been proven to be beneficial for learning (see, Letón, Molanes-López, Luque, & Conejo, 2018; Luzón & Letón, 2015; Rodríguez-Ascaso, Letón, Muñoz-Carenas, & Finat, 2018). The question for each MTM is randomly assigned by the teacher from a database of true/false questions. The pedagogical values of online asynchronous discussions in developing learners' reflective capabilities and critical thinking skills (Barnett-Queen, Blair, & Merrick, 2005; Whipp, 2003) are known. However, it has been reported that forced participation can cause anxiety, resistance, and resentment among students (Pena-Shaff, Altman, & Stephenson, 2005). To facilitate socialisation during the course, a series of activities are performed following an approach similar to that of Deng and Tavares (2013): a daily true/false question and a voluntary mock exam on Twitter.

## 4.2 | Interaction data analysis

### 4.2.1 | Interaction data retrieval

We introduced social factors into the reviewer assignment method using the learners' interactions on the forums and analysed these interactions with SNA. The forums are the main means of communication among those involved in the course, so the social interactions considered were those conducted through the forums. We extracted communications in the form of forum messages to determine who had sent each message and to whom each message had been sent. In this research, we did not use any personal information about the learners. We anonymised the forum interactions before performing any analysis.

### 4.2.2 | Interaction data analysis process

Our proposed variables can be divided into statistical variables and SNA variables. For the statistical analysis, we counted the number of conversations initiated by each learner to calculate the variable Initiative and the number of messages that each learner sent to each other learner to calculate the variable SharedMessages. To analyse the social relationships between the learners, we built a table in which each row showed, which learner had sent messages to other learners, and each column showed, which learner had received messages from other learners. With this table, we were able to perform SNA using the Gephi tool.[1]

Thus, we calculated the network attributes described in Section 3.1. Figure 4 shows an example of an SNA metric, in this case, the learners' eigenvector centrality. The reddest nodes are those with the highest values of the eigenvector centrality. Each node represents a different (anonymised) learner.

Once the SNA was complete, we created a data set consisting of the SNA variables and the variable Initiative to feed into the ID. This data set is called the *reviewer features* data set and is shown in Table 4. We note that the values of the numeric variables are normalised and that the variable StudentMessages is discretised into three values: high, medium, and low. We obtained the values of this variable from the SNA metric Outdegree, which we also used to infer the probabilistic variable Activity (see Sections 3.1 and 3.3). In this data set, the shared messages

[1]https://gephi.org

**TABLE 4** Social network analysis results

| Learner | Initiative | Popularity | Closeness | Betweenness | Eigenvector | StudentMessages |
|---|---|---|---|---|---|---|
| L-1 | 0.166666667 | 0.4 | 0.380952381 | 0.03448 | 0.46776156 | High |
| L-2 | 0.166666667 | 0.4 | 0.41025641 | 0.04260 | 0.322054925 | High |
| L-3 | 0.666666667 | 1 | 0 | 0.00000 | 0.356993916 | Medium |
| L-4 | 0 | 0 | 0.346938776 | 0.00000 | 0 | High |
| L-5 | 0.166666667 | 0.4 | 0 | 0.00000 | 0.354758766 | Low |
| L-6 | 0 | 0 | 1 | 0.00000 | 0 | Medium |
| L-7 | 0 | 0 | 1 | 0.00000 | 0 | Medium |
| L-8 | 0.333333333 | 0.2 | 0.352941176 | 0.00000 | 0.096066541 | Medium |
| L-9 | 0.166666667 | 0.2 | 0.666666667 | 0.00000 | 0.012553811 | Medium |
| L-10 | 0 | 0 | 1 | 0.00000 | 0 | Medium |
| L-11 | 0 | 0.2 | 1 | 0.01217 | 0.012553811 | High |
| L-12 | 0.5 | 1 | 0.4 | 0.63286 | 0.644024067 | High |
| L-13 | 0 | 0 | 0 | 0.00000 | 0 | Low |
| L-14 | 0.166666667 | 0.6 | 0.666666667 | 0.47465 | 0.30673291 | High |
| L-15 | 0 | 0.2 | 0 | 0.00000 | 0.320464518 | Low |
| L-16 | 0 | 0 | 0.391304348 | 0.00000 | 0 | High |
| L-17 | 0.166666667 | 0.8 | 0 | 0.00000 | 0.129475767 | Low |
| L-18 | 0.166666667 | 0.8 | 0 | 0.00000 | 0.118959385 | Low |
| L-19 | 0 | 0.2 | 1 | 0.01217 | 0.003204645 | Medium |
| L-20 | 0 | 0 | 0 | 0.00000 | 0 | Low |
| L-21 | 0 | 0.6 | 0.6 | 0.61055 | 0.639993483 | High |
| L-22 | 0.5 | 0.8 | 0.432432432 | 0.16227 | 0.643271755 | High |
| L-23 | 0 | 0.4 | 0.296296296 | 0.00000 | 0.429883595 | Medium |
| L-24 | 1 | 0.6 | 0.5 | 0.19473 | 0.245358993 | High |
| L-25 | 0.333333333 | 0.6 | 1 | 0.10953 | 0.326873809 | Medium |
| L-26 | 0.333333333 | 1 | 0.551724138 | 1.00000 | 1 | High |
| L-27 | 0.166666667 | 0 | 0 | 0.00000 | 0 | Low |
| L-28 | 0 | 0 | 1 | 0.00000 | 0 | Medium |
| L-29 | 0 | 0.2 | 0.290909091 | 0.00000 | 0.212635848 | Medium |
| L-30 | 0 | 0 | 0.666666667 | 0.00000 | 0 | Medium |
| L-31 | 0 | 0 | 1 | 0.00000 | 0 | Medium |

exchanged between two learners are not displayed. We built another data set consisting of the numbers of messages exchanged by each learner with each other learner. Because the number of learners was 31, these data comprised $31 \times 30 = 930$ instances. This data set is called the *shared messages* data set. Both data sets, *reviewer features* and *shared messages*, were used when building the ID.

## 4.3 | Completing the elicitation of the numerical parameters of the influence diagram with the SNA results

In Section 3.3.2, we described how to elicit the numerical parameters of the model for the course-independent ID template. In this section, we explain how we elicited the rest of the numerical parameters from the SNA results, thereby obtaining an ID network.

The probability table of the variable StudentMessages was obtained by discretising the set of the number of messages sent by each student into three levels (or states), low, medium, and high, using an equal-frequency binning method and then assigning the corresponding label to each student. Thus, the probability table of this variable is deterministic; for each student, the probability of StudentMessages is 1.0 for the state corresponding to the assigned label and 0.0 for the other two states. The probability table of ReviewerMessages is identical to that of StudentMessages, except that Reviewer, rather than Student, is the conditioning variable.

## 4.4 | Student assignment process

The main objective of the model we built is to find a method of selecting students for an OPA task that can consider social factors. One of the main outputs of an ID evaluation is a policy, in the form of a table, for each decision *D* in the model that indicates, for each configuration of values of the known variables affecting *D*, which is the best option to select for *D*. Thus, the ID evaluation produces the information that we want to obtain: given a student, we wish to select which coursemate is the best choice for reviewing his/her exercise.

**TABLE 5** Peer-to-peer assignment using the weighting function $f(x) = x^2$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. L-1 | | 25 | 21 | 21 | 12 | 29 | 34 | 22 | 25 | 34 | 33 | 34 | 9 | 38 | 10 | 22 | 12 | 12 | 29 | 9 | 31 | 31 | 18 | 37 | 35 | 49 | 11 | 29 | 17 | 23 | 29 |
| 2. L-2 | 25 | | 21 | 21 | 12 | 29 | 29 | 22 | 25 | 29 | 33 | 50 | 9 | 33 | 10 | 22 | 12 | 12 | 29 | 9 | 31 | 42 | 23 | 37 | 35 | 43 | 11 | 29 | 17 | 23 | 29 |
| 3. L-3 | 21 | 21 | | 22 | 14 | 33 | 33 | 25 | 29 | 33 | 29 | 30 | 11 | 29 | 12 | 18 | 13 | 13 | 38 | 11 | 33 | 27 | 22 | 33 | 39 | 34 | 12 | 38 | 21 | 27 | 33 |
| 4. L-4 | 25 | 25 | 26 | | 12 | 29 | 29 | 22 | 25 | 29 | 33 | 39 | 9 | 33 | 10 | 22 | 17 | 12 | 29 | 9 | 31 | 31 | 18 | 37 | 40 | 38 | 11 | 29 | 17 | 23 | 29 |
| 5. L-5 | 19 | 20 | 21 | 15 | | 29 | 29 | 22 | 25 | 29 | 27 | 28 | 15 | 27 | 16 | 16 | 17 | 17 | 29 | 15 | 26 | 26 | 18 | 32 | 35 | 33 | 16 | 29 | 17 | 23 | 29 |
| 6. L-6 | 21 | 21 | 24 | 17 | 14 | | 33 | 25 | 29 | 33 | 29 | 30 | 11 | 29 | 12 | 18 | 19 | 13 | 33 | 11 | 27 | 27 | 22 | 33 | 39 | 34 | 12 | 33 | 21 | 27 | 33 |
| 7. L-7 | 26 | 21 | 24 | 17 | 14 | 33 | | 25 | 29 | 33 | 29 | 30 | 11 | 29 | 12 | 18 | 13 | 13 | 33 | 11 | 27 | 27 | 22 | 33 | 39 | 34 | 12 | 33 | 21 | 27 | 33 |
| 8. L-8 | 21 | 21 | 24 | 17 | 14 | 33 | 33 | | 29 | 33 | 29 | 30 | 11 | 29 | 12 | 18 | 13 | 13 | 33 | 11 | 27 | 27 | 22 | 54 | 39 | 34 | 12 | 33 | 21 | 27 | 33 |
| 9. L-9 | 21 | 21 | 24 | 17 | 14 | 33 | 33 | 25 | | 33 | 39 | 30 | 11 | 29 | 12 | 18 | 13 | 13 | 33 | 11 | 27 | 27 | 22 | 33 | 39 | 34 | 12 | 33 | 21 | 27 | 33 |
| 10. L-10 | 26 | 21 | 24 | 17 | 14 | 33 | 33 | 25 | 29 | | 29 | 30 | 11 | 29 | 12 | 18 | 13 | 13 | 33 | 11 | 27 | 27 | 22 | 33 | 39 | 34 | 12 | 33 | 21 | 27 | 33 |
| 11. L-11 | 25 | 25 | 21 | 21 | 12 | 29 | 29 | 22 | 36 | 29 | | 34 | 9 | 33 | 10 | 22 | 17 | 12 | 29 | 9 | 31 | 31 | 18 | 37 | 35 | 38 | 11 | 29 | 17 | 23 | 29 |
| 12. L-12 | 25 | 46 | 21 | 26 | 12 | 29 | 29 | 22 | 25 | 29 | 33 | | 9 | 33 | 10 | 22 | 12 | 12 | 29 | 9 | 31 | 31 | 34 | 37 | 35 | 64 | 11 | 29 | 28 | 23 | 29 |
| 13. L-13 | 19 | 20 | 21 | 15 | 18 | 29 | 29 | 22 | 25 | 29 | 27 | 28 | | 27 | 16 | 16 | 17 | 17 | 29 | 15 | 26 | 26 | 18 | 32 | 35 | 33 | 16 | 29 | 17 | 23 | 29 |
| 14. L-14 | 30 | 25 | 21 | 21 | 12 | 29 | 29 | 22 | 25 | 29 | 33 | 34 | 9 | | 10 | 27 | 12 | 12 | 29 | 9 | 42 | 31 | 18 | 48 | 35 | 38 | 11 | 29 | 17 | 23 | 29 |
| 15. L-15 | 19 | 20 | 21 | 15 | 18 | 29 | 29 | 22 | 25 | 29 | 27 | 28 | 15 | 27 | | 16 | 17 | 17 | 29 | 15 | 26 | 26 | 18 | 32 | 35 | 38 | 16 | 29 | 17 | 23 | 29 |
| 16. L-16 | 25 | 25 | 21 | 21 | 12 | 29 | 29 | 22 | 25 | 29 | 33 | 34 | 9 | 38 | 10 | | 12 | 12 | 35 | 9 | 31 | 37 | 18 | 37 | 35 | 38 | 11 | 29 | 17 | 23 | 29 |
| 17. L-17 | 19 | 20 | 21 | 20 | 18 | 34 | 29 | 22 | 25 | 29 | 32 | 28 | 15 | 27 | 16 | 16 | | 17 | 29 | 15 | 26 | 26 | 18 | 32 | 40 | 33 | 16 | 29 | 17 | 23 | 29 |
| 18. L-18 | 19 | 20 | 21 | 15 | 18 | 29 | 29 | 22 | 25 | 29 | 27 | 28 | 15 | 27 | 16 | 16 | 17 | | 29 | 15 | 26 | 26 | 18 | 32 | 35 | 33 | 16 | 29 | 17 | 23 | 34 |
| 19. L-19 | 21 | 21 | 29 | 17 | 14 | 33 | 33 | 25 | 29 | 33 | 29 | 30 | 11 | 29 | 12 | 23 | 13 | 13 | | 11 | 27 | 27 | 22 | 33 | 39 | 34 | 12 | 33 | 21 | 27 | 33 |
| 20. L-20 | 19 | 20 | 21 | 15 | 18 | 29 | 29 | 22 | 25 | 29 | 27 | 28 | 15 | 27 | 16 | 16 | 17 | 17 | 29 | | 26 | 26 | 18 | 32 | 35 | 33 | 16 | 29 | 17 | 23 | 29 |
| 21. L-21 | 25 | 25 | 26 | 21 | 12 | 29 | 29 | 22 | 25 | 29 | 33 | 34 | 9 | 43 | 10 | 22 | 12 | 12 | 29 | 9 | | 37 | 18 | 37 | 35 | 43 | 11 | 29 | 17 | 23 | 29 |
| 22. L-22 | 25 | 36 | 21 | 21 | 12 | 29 | 29 | 22 | 25 | 29 | 33 | 34 | 9 | 33 | 10 | 27 | 12 | 12 | 29 | 9 | 37 | | 18 | 37 | 35 | 59 | 11 | 29 | 17 | 23 | 29 |
| 23. L-23 | 21 | 26 | 24 | 17 | 14 | 33 | 33 | 25 | 29 | 33 | 29 | 35 | 11 | 29 | 12 | 18 | 13 | 13 | 33 | 11 | 27 | 27 | | 33 | 39 | 39 | 12 | 33 | 21 | 27 | 33 |
| 24. L-24 | 25 | 25 | 21 | 21 | 12 | 29 | 29 | 43 | 25 | 29 | 33 | 34 | 9 | 43 | 10 | 22 | 12 | 12 | 29 | 9 | 31 | 31 | 18 | | 35 | 38 | 11 | 29 | 17 | 23 | 29 |
| 25. L-25 | 21 | 21 | 24 | 22 | 14 | 33 | 33 | 25 | 29 | 33 | 29 | 30 | 11 | 29 | 12 | 18 | 19 | 13 | 33 | 11 | 27 | 27 | 22 | 33 | | 39 | 12 | 33 | 21 | 32 | 33 |
| 26. L-26 | 35 | 30 | 21 | 21 | 12 | 29 | 29 | 22 | 25 | 29 | 33 | 60 | 9 | 33 | 15 | 22 | 12 | 12 | 29 | 9 | 37 | 52 | 23 | 37 | 40 | | 11 | 29 | 17 | 23 | 29 |
| 27. L-27 | 19 | 20 | 21 | 15 | 18 | 29 | 29 | 22 | 25 | 29 | 27 | 28 | 15 | 27 | 16 | 16 | 17 | 17 | 29 | 15 | 26 | 26 | 18 | 32 | 35 | 33 | | 29 | 17 | 23 | 29 |
| 28. L-28 | 21 | 21 | 29 | 17 | 14 | 33 | 33 | 25 | 29 | 33 | 29 | 30 | 11 | 29 | 12 | 18 | 13 | 13 | 33 | 11 | 27 | 27 | 22 | 33 | 39 | 34 | 12 | | 21 | 27 | 33 |
| 29. L-29 | 21 | 21 | 24 | 17 | 14 | 33 | 33 | 25 | 29 | 33 | 29 | 40 | 11 | 29 | 12 | 18 | 13 | 13 | 33 | 11 | 27 | 27 | 22 | 33 | 39 | 34 | 12 | 33 | | 27 | 33 |
| 30. L-30 | 21 | 21 | 24 | 17 | 14 | 33 | 33 | 25 | 29 | 33 | 29 | 30 | 11 | 29 | 12 | 18 | 13 | 13 | 33 | 11 | 27 | 27 | 22 | 33 | 44 | 34 | 12 | 33 | 21 | | 33 |
| 31. L-31 | 21 | 21 | 24 | 17 | 14 | 33 | 33 | 25 | 29 | 33 | 29 | 30 | 11 | 29 | 12 | 18 | 13 | 18 | 33 | 11 | 27 | 27 | 22 | 33 | 39 | 34 | 12 | 33 | 21 | 27 | |

*Note.* The color coding scheme is explained in Section 4.5.2.

We designed and implemented an ID-network-based algorithm for finding such an assignment. The implementation was divided into two stages.

In the first stage, we calculated $EU(s, r)$, a table that, based on the ID network, gives the expected utility of selecting reviewer $r$ to review the exercise of student $s$. This table was calculated with OpenMarkov (Arias & Díez, 2008), an open-source tool for probabilistic graphical models, which is available at www.openmarkov.org.

Table 5 displays $EU(s, r)$ for the set of all students in the course, with the values normalised to the interval $(0, 100)$. The cell in row $i$ and column $j$ indicates the expected utility of selecting student $j$ to act as a reviewer for $i$.

In the second stage, we implemented an algorithm for assigning reviewers to each student based on the table of utilities given by $EU(s, r)$. The algorithm first assigns one reviewer to each student in the course, then another reviewer to each student, and so on, until it has assigned $n$ reviewers to each student. For an objective review process, ideally, $n$ should be greater than 1. In Table 5, we have used $n = 3$ because it is hard to imagine that asking students to conduct more than three peer reviews would boost the acceptance rate, which is in accordance with the work by Sung, Chang, Chang, and Yu (2010).

In each iteration of the algorithm, given a student $x$, it first calculates the set of reviewer candidates for $x$ under the constraints that a student cannot review himself/herself and that all reviewers for a student must be different. From the set of possible reviewers $\mathbf{S}$ that can be assigned to $x$, the algorithm then selects the reviewer $r$ for whom $r = argmax_{y \in \mathbf{S}} EU(x, y)$ is satisfied.

Finally, after the algorithm has returned an assignment of $n$ reviewers for each student in $\mathbf{S}$, we can measure the quality of the overall assignment as follows:

$$Quality = \frac{\sum_{s \in \mathbf{S}} \sum_{r \in reviewers(s)} EU(s, r)}{n \cdot |\mathbf{S}|}, \tag{2}$$

where $|\mathbf{S}|$ is the cardinal of $\mathbf{S}$ and reviewers(s) is the set of reviewers assigned to student $s$.

We argue that the higher the quality value of an assignment is, the more appropriate it is for the OPA task.

## 4.5 | Experiments on computing assignments

The algorithm described in Section 4.4 is nondeterministic because in each iteration, it randomly selects a student and assigns a reviewer to him or her. Thus, different executions can provide different assignments of different quality. We ran the above algorithm 10,000 times and then selected the assignment with the highest quality value, as calculated using Equation (2), as the best one.

### 4.5.1 | Comparison of the expected utility-based algorithm with random selection

To assess the proposed algorithm, we also implemented and executed a random algorithm that did not use the ID information (the table of utilities $EU(s, r)$) to select reviewers but instead randomly selected each reviewer from the set of possible candidates.

The results of comparing the expected utility-based algorithm proposed in Section 4.4 with the random selection algorithm are presented in Table 6. The mean and standard deviation of the quality of the best assignment found after 10,000 executions of each algorithm, as measured using Equation (2), are presented in that table. By examining data summarised in Table 6 through an ANOVA one-way with two groups (Random vs. EU Maximising), it can be seen that in all three weighting functions, there are statistical significant differences between algorithms ($p$ value $< 0.0001$), so we can conclude that the proposed algorithm outperforms the random algorithm.

For each weighting function, we also calculated the two metrics presented in the sixth and seventh columns of Table 6:

1. The theoretical bound is the quality of the assignment that we would obtain if the best $n$ reviewers for each student could be selected by allowing a student to review more than $n$ students (or fewer than $n$ students). Let $Q^*$ denote the quality value of this new assignment obtained after 10,000 executions. The value of $Q^*$ depends on the value of $n$, as the value of $Q^*$ decreases when the value of $n$ increases. We must try to find a balance between having a low value of $n$ (and high value of $Q^*$) and ensuring that each student has been reviewed by at least one student. We have used $n = 3$, as we mentioned previously in this section.
2. The percentage improvement achieved with our proposed algorithm can be calculated as follows:

$$\frac{Quality_{EUMax} - Quality_{Random}}{Q^* - Quality_{Random}} \times 100,$$

where $Quality_{EUMax}$ is the mean quality achieved with the proposed algorithm (EU Maximising) and $Quality_{Random}$ is the mean value of the quality achieved with the random algorithm.

For example, when using the weighting function $f(x) = x^2$, we obtained a theoretical quality bound of $Q^* = 38.182$. In Table 6, we can see that $Quality_{EUMax} = 27.206$ and $Quality_{Random} = 24.206$. Thus, the percentage improvement with our algorithm was $(27.206 - 24.206)/(38.182 - 24.206) \times 100 = 20\,\%$.

**TABLE 6** Results obtained using different weighting functions

| Algorithm | Random | | EU Maximising | | | |
|---|---|---|---|---|---|---|
| Weighting function $f(x)$ | Mean | SD | Mean | SD | Theoretical bound | Improvement with 'Max. EU' |
| $x$ | 23.792714993855586 | 0.2619577679093355 | 25.359546839093603 | 0.3047333245345558 | 42.0174668970814 | 9 % |
| $x^2$ | 24.473428779800294 | 0.3250034734662528 | 27.206333478341016 | 0.3610816683908969 | 38.18154956620584 | 20 % |
| $\sqrt{x}$ | 23.632757146307977 | 0.21544897697694051 | 24.761501206426022 | 0.24958485139279954 | 45.961087511631604 | 5 % |

*Note*. SD: standard deviation.

### 4.5.2 | Computation of the best assignment

We tested three weighting functions $f(x)$, as described in Section 3.3.2: $f(x) = x$, $f(x) = x^2$, and $f(x) = \sqrt{x}$. For each weighting function, we identified the assignment with the best quality and calculated the number of cases in which an optimal reviewer was selected. This calculation can be illustrated by considering Table 5, in which the cells are coloured as follows. Let $A$ be the best assignment calculated with our algorithm, and let $A^*$ be the assignment that achieves the theoretical quality bound $Q^*$ as described in the previous subsection. In Table 5, the colour code given below is used to indicate when the student in the $j$ th column is selected as a reviewer for the student in the $i$ th row:

- Yellow: Selected in $A^*$, but not in $A$.
- Blue: Selected in $A$, but not in $A^*$.
- Green: Selected in $A$ and $A^*$.
- White: Not selected in $A$ or $A^*$.

The third column of Table 7 presents the number of cases in which an optimal reviewer was selected in $A$. Thus, it corresponds to the number of green cells in an assignment table such as Table 5. The last column of Table 7 presents the percentage of cases in which an optimal reviewer was selected.

## 5 | EVALUATION OF OUR METHOD

We have proposed a method of assigning reviewers for an OPA task in a real education environment. Our main objective was to encourage learners to participate in the OPA task to take advantage of the pedagogical benefits of this kind of learning task (Deng & Tavares, 2013; Uto & Ueno, 2016), and our hypothesis was that learners consider social relations when participating in an OPA task. We performed a simple experiment to check our hypothesis.

### 5.1 | Experiment

One hundred learners were enrolled in the course introduced in the previous section. Throughout the 14 weeks of the course, 45 of these 100 learners participated in the LMS, and 51 of the 100 learners recorded MTMs. Although these percentages of participation in continuous assessment are not very high, they are very good when compared with the 66% of the learners who took the final evaluation.

The OPA task was scheduled in week #5, when the true/false questions to be solved in the MTMs were assigned. The learners had until week #13 to upload their final MTMs to YouTube and to share the URLs on the LMS so that everyone could see them. During this time, the teachers encouraged the learners to participate in the LMS by answering theoretical questions in MTMs. Additionally, to increase participation in the LMS, the teachers conducted a session with simple true/false questions and a mock exam on Twitter and a mock exam based on true/false questions similar to those posed for the MTMs was also presented on the LMS. In week #14, the real OPA took place. For the OPA task, each learner was asked by personal email to assess three MTMs during the last week, in the same line as Sung et al. (2010). The learners were provided with a rubric with which to assess each MTM. The highest possible score on the rubric consisted of 6 points for the proper presentation of the MTM (0.5 points for each of the 12 items used to assess the degree of fulfilment of the MTM features described by ; Letón & Molanes-López, 2014) and 4 points for providing the correct answer in the MTM. Notably, 6 points were given if the MTM was properly done, even if the answer was incorrect. This is because we wanted to assess other abilities in addition to academic ones.

Our hypothesis was as follows:

H1: Considering social factors when assigning reviewers improves participation in OPA tasks.

To validate our hypothesis, we designed an experiment to verify our social analysis-based reviewer assignment method. We divided the learners who recorded MTMs into two groups: an *experimental group*, consisting of learners whose forum interactions were analysed and a *control group*, consisting of learners whose forum interactions were not analysed. We recorded the participation of both learner groups in the OPA task.

**TABLE 7** Metrics for the best assignment obtained for each weighting function

| Weighting function $f(x) :=$ | Best expected utility | Number of cases of optimal selection | Percentage of the cases (%) |
|---|---|---|---|
| $x$ | 26.397183811059904 | 16 | 17 |
| $x^2$ | 28.628737717972353 | 24 | 25 |
| $\sqrt{x}$ | 25.711613163400386 | 13 | 13 |

**TABLE 8** Experimental design and results

| Cluster | # Requests in experimental group | #Requests in control group |
|---|---|---|
| 3 | 66 (70.97%) | 30 (50.00%) |
| 2 | 4 (4.30%) | 6 (10.00%) |
| 1 | 1 (1.08%) | 0 (0.00%) |
| 0 | 22 (23.66%) | 24 (40.00%) |
| Total | 93 | 60 |

## 5.2 | Results

There were 31 learners in the experimental group, each of whom was able to be reviewed by three other learners in the same group. The reviewers were selected from among the learners in this group using the reviewer assignment method explained in this paper. Then, three requests for review were sent to each learner. The students were not obliged to accept all the requests, but they could accept only some of them. Of these 93 requests, 71 were accepted. Thus, the percentage of acceptance was 76.34%.

There were 20 learners in the control group, each of whom was similarly reviewed by three other learners in the same group. The reviewers were selected randomly from among the learners in this group. Then, three requests for review were sent to each learner. Of these 60 requests, 36 were accepted. The percentage of acceptance was 60.00%.

As summarised in Table 8, there are four clusters: Cluster *i*, where *i* is an integer, $0 < i \leq 3$, corresponds to the requests accepted by students that accepted exactly *i* requests; and Cluster 0 corresponds to the remaining requests. There is statistically significant dependence among the clusters in the two groups ($\chi^2_3 = 8.2533$, $p$ value $= 0.0411$). This finding suggests that our reviewer assignment method improves participation in OPA tasks. Because social factors were considered to improve the assignment process, we can deduce that social factors encourage participation in OPA tasks in online courses. Therefore, our experimental results support our hypothesis. It is important to note that the participation in the experimental group was 16 percentage points higher than that in the control group if we consider Clusters 3, 2, and 1, and 21 percentage higher if we only consider Cluster 3. These differences are considered relevant in terms of practical importance to improve the acceptance rate, although more work is needed to see measurable improvements in learning outcomes.

# 6 | CONCLUSIONS AND FUTURE WORKS

We have proposed a social analysis-based reviewer assignment method for use in OPA tasks. The template for the method is independent of any specific course; we used it in an implementation for a real OPA task, the results of which support that our approach improves participation. Our proposed assignment method encouraged learners to participate in the OPA task, resulting in a participation percentage of 76.34% compared with the 60.00% participation observed with random assignment.

We would like to emphasise that our social analysis-based reviewer assignment method can be easily adapted to different criteria by changing the numerical parameters in the ID template. We observe that the assignment criteria for an OPA task can vary depending on the educational context or the experience of the teachers. In addition, we note that other SNA metrics could be used depending on the context. At present, these changes must be made by an expert in data mining or Bayesian theory. Consequently, we will need to improve our assignment method to allow a non-expert to easily adapt it to a new context.

We tested our social analysis-based reviewer assignment method in one real educational environment. To study the advantages of our assignment method in greater depth, we will need to test it in other learning environments that do not share the same educational context. For this purpose, we will need to work with other experts to improve the simplicity and flexibility of our assignment method.

Because of the size of our network, which we analysed with SNA techniques, we did not study the network structure and the students segmentation. To clarify the strengths and weaknesses of our approach, we are planning a future experience, with more students, to take into account the network structure and its evolution, to measure improvements in learning outcomes and to study the reliability (consistency of scores given by multiple reviewers) and validity (correlation between reviewer-assigned scores and teacher-assigned scores) in our social analysis-based reviewer assignment method.

In addition, the algorithm proposed in this paper has a very high computational complexity when applied to a course with a large number of students, such as a MOOC. In such a case, finding the best assignment may require a very high number of executions. We would like to investigate heuristic methods that can accelerate the search for the best assignment to reduce the computational time required.

Some of the assumptions of our method can be relaxed to adapt it to more complex and dynamic situations encountered in online courses. First, our algorithm assigns reviewers only after every student has performed the exercise. However, the algorithm could assign a reviewer for a student's work immediately after the exercise has been completed. Second, our algorithm assumes that each reviewer will eventually perform the assigned reviews. Unfortunately, that assumption is not realistic in online courses, in which the dropout rate is usually high. It is reasonable to suppose that students who have dropped out of the course will typically not want to continue as reviewers. Therefore, the algorithm should be able to find a new reviewer if the initially selected reviewer ceases to participate in the course at a later date.

As a related concern, we would like to research whether social interactions in conjunction with OPA can reduce the dropout rate in a distance education environment.

## CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

## ORCID

*Antonio R. Anaya* https://orcid.org/0000-0002-3336-7082
*Manuel Luque* https://orcid.org/0000-0003-3018-3760
*Emilio Letón* https://orcid.org/0000-0001-6908-9590
*Félix Hernández-del-Olmo* https://orcid.org/0000-0002-0567-9572

## REFERENCES

Aggarwal, C. C. (2011). An introduction to social network data analytics. In Aggarwal, C. C. (Ed.), *Social network data analytics*. Boston, MA: Springer US, pp. 1–15.

Anaya, A. R., & Boticario, J. G. (2011). Content-free collaborative learning modeling using data mining. *User Modeling and User-adapted Interaction*, *21*, 181–216.

Anaya, A. R., Luque, M., & García-Saiz, T. (2013). Recommender system in collaborative learning environment using an influence diagram. *Expert Systems with Applications*, *40*, 7193–7202.

Arias, M., & Díez, F. J. (2008). Carmen: An open source project for probabilistic graphical models. In *Proceedings of the Fourth European Workshop on Probabilistic Graphical Models (PGM'08)*, Hirtshals, Denmark, pp. 25–32.

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education*, *27*, 427–441.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of educational research*, *61*, 213–238.

Barnett-Queen, T., Blair, R., & Merrick, M. (2005). Student perspectives of online discussions: Strengths and weaknesses. *Journal of Technology in Human Services*, *23*, 229–244.

Bouguessa, M., & Romdhane, L. B. (2015). Identifying authorities in online communities. *ACM Transactions on Intelligent Systems and Technology*, *6*, 1–23.

Bouzidi, L., & Jaillet, A. (2009). Can online peer assessment be trusted? *Journal of Educational Technology & Society*, *12*, 257–268.

Capuano, N., & Caballé, S. (2015). Towards adaptive peer assessment for MOOCs. In *2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)*, Krakow, Poland, pp. 64–69.

Cheung, W. S., Hew, K. F., & Ng, C. S. L. (2008). Toward an understanding of why students contribute in asynchronous online discussions. *Journal of Educational Computing Research*, *38*, 29–50.

Daradoumis, T., Juan, A. A., Lera-López, F., & Faulin, J. (2010). Using collaboration strategies to support the monitoring of online collaborative learning activity, *Technology Enhanced Learning. Quality of Teaching and Educational Reform*. Berlin, Germany: Springer Verlag, pp. 271–277.

Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Educational Technology and Society*, *11*, 224–238.

Deng, L., & Tavares, N. J. (2013). From Moodle to Facebook: Exploring students' motivation and experiences in online communities. *Communication Education*, *68*, 167–176.

Engestrom, Y. (2000). Activity theory as a framework for analyzing and redesigning work. *Ergonomics*, *43*, 960–974.

Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment & Evaluation in Higher Education*, *11*, 146–166.

Gielen, M., & De Wever, B. (2015). Scripting the role of assessor and assessee in peer assessment in a wiki environment: Impact on peer feedback quality and product improvement. *Computers & Education*, *88*, 370–386.

Hanrahan, S. J., & Isaacs, G. (2001). Assessing self-and peer-assessment: The students' views. *Higher Education Research & Development*, *20*, 53–70.

Hopcroft, J. E., Motwani, R., & Ullman, J. D. (2007). *Teoría de autómatas, lenguajes y computación*. Madrid, Spain: Addison-Wesley.

Hoskins, S. L., & Van Hooff, J. C. (2005). Motivation and ability: Which students use online learning and what influence does it have on their achievement? *British Journal of Educational Technology*, *36*, 177–192.

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education*, *71*, 133–152.

Howard, R. A., & Matheson, J. E. (1984). Influence diagrams. In Howard, R. A., & Matheson, J. E. (Eds.), *Readings on the principles and applications of decision analysis*. Menlo Park, CA: Strategic Decisions Group, pp. 719–762.

Huang, E. Y., Lin, S. W., & Huang, T. K. (2012). What type of learning style leads to online participation in the mixed-mode e-learning environment? A study of software usage instruction. *Communication Education*, *58*, 338–349.

Johnson, D. W., & Johnson, R. T. (2004). Cooperation and the use of technology, *Handbook of research on educational communications and technology*: Taylor & Francis, pp. 785–812.

Johnson, D. W., & Johnson, R. T. (2005). Learning groups. In Wheelan, S. A. (Ed.), *The handbook of group research and practice*. Thousand Oaks, CA: Sage, pp. 441–461.

Jonassen, D. H., Peck, K. L., & Wilson, B. G. (1999). *Learning with technology: A constructivist perspective*. Upper Saddle River, N.J.: Merrill/Prentice-Hall.

Kao, G. Y.-M. (2013). Enhancing the quality of peer review by reducing student "free riding": Peer assessment with positive interdependence. *British Journal of Educational Technology*, *44*, 112–124.

Lacave, C., Luque, M., & Díez, F. J. (2007). Explanation of Bayesian networks and influence diagrams in Elvira. *IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics*, *37*, 952–965.

Letón, E., & Molanes-López, E. M. (2014). Two new concepts in video podcasts-Minimalist Slides and Modular Teaching Mini-videos. In *Proceedings of the 6th International Conference on Computer Supported Education (CSEDU'2014)*, Barcelona, Spain, pp. 292–297.

Letón, E., Molanes-López, E. M., Luque, M., & Conejo, R. (2018). Video podcast and illustrated text feedback in a web-based formative assessment environment. *Computer Applications in Engineering Education*, *26*, 187–202.

Lu, Y., Warren, J., Jermaine, C., Chaudhuri, S., & Rixner, S. (2015). Grading the graders: Motivating peer graders in a MOOC. In *Proceedings of the 24th International Conference on World Wide Web*, Florence, Italy, pp. 680–690.

Luque, M., & Díez, F. J. (2010). Variable elimination for influence diagrams with super-value nodes. *International Journal of Approximate Reasoning*, *51*, 615–631.

Luzón, J. M., & Letón, E. (2015). Use of animated text to improve the learning of basic mathematics. *Computers & Education*, *88*, 119–128.

Meek, S. E. M., Blakemore, L., & Marks, L. (2016). Is peer review an appropriate form of assessment in a MOOC? Student participation and performance in formative peer review. *Assessment & Evaluation in Higher Education*, *42*, 1000–1013.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.

Pena-Shaff, J., Altman, W., & Stephenson, H. (2005). Asynchronous online discussions as a tool for learning: Students' attitudes, expectations, and perceptions. *Journal of Interactive Learning Research*, *16*, 409–431.

Rodríguez-Ascaso, A., Letón, E., Muñoz-Carenas, J., & Finat, C. (2018). Accessible mathematics videos for non-disabled students in primary education. *PLoS One*, *13*, 1–18.

Rosa, S. S., Coutinho, C. P., & Flores, M. A. (2016). Online peer assessment: Method and digital technologies. *Procedia-Social and Behavioral Sciences*, *228*, 418–423.

Sagheb-Tehrani, M. (2006). The design process of expert systems development: Some concerns. *Expert Systems*, *23*, 116–125.

Scott, J. (2017). *Social network analysis*. London: Sage.

Sinha, T. (2014). Together we stand, together we fall, together we win: Dynamic team formation in massive open online courses. In *2014 Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT)*, IEEE, India, pp. 107–112.

Sung, Y.-T., Chang, K.-E., Chang, T.-H., & Yu, W.-C. (2010). How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments. *Journal of Adolescence*, *33*, 135–145.

Tatman, J. A., & Shachter, R. D. (1990). Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man and Cybernetics*, *20*, 365–379.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*, 249–276.

Topping, K. J. (2005). Trends in peer learning. *Educational psychology*, *25*, 631–645.

Ueno, M., & Okamoto, T. (2008). Item response theory for peer assessment. In *2008 Eighth IEEE International Conference on Advanced Learning Technologies*, Santander, Cantabria, Spain, pp. 554–558.

Uto, M., & Ueno, M. (2016). Item response theory for peer assessment. *IEEE Transactions On Learning Technologies*, *9*, 157–170.

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, *4*, 41–54.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and Applications*, Vol. *8*. Cambridge: Cambridge University Press.

Whipp, J. L. (2003). Scaffolding critical reflection in online discussions helping prospective teachers think deeply about field experiences in urban schools. *Journal of Teacher Education*, *54*, 321–333.

Zafarani, R., Abbasi, M. A., & Liu, H. (2014). *Social media mining: An introduction*. New York, NY: Cambridge University Press.

## AUTHOR BIOGRAPHIES

**Antonio R. Anaya** was born in Madrid, Spain, in 1974. He is Bachelor of Astrophysics, Universidad Complutense de Madrid, since 1999. He got the PhD in Artificial Intelligence in UNED (National University of Distance Education of Spain) in 2009. He works as Assistant Professor since 2014 until now. His most important research has focused on the data mining techniques in the context of e-learning collaborative environments and stellar evolution using artificial intelligence tools.

**Manuel Luque** was born in Madrid, Spain, in 1977. He received a Computer Science bachelor's degree from the University of Málaga, Spain, in 2003, and a PhD degree from the Universidad Nacional de Educación a Distancia (UNED), Madrid, in 2009. He is currently an Associate Professor in the Department of Artificial Intelligence at UNED, where he has worked since 2003. In addition, he is a member of the Research Center for Intelligent Decision-Support Systems (CISIAD) at the same university. His research interests focus on the theory, methods, and applications of probabilistic graphical models, especially in the fields of medicine and education. Personal web page: www.ia.uned.es/~mluque.

**Emilio Letón** is born in Madrid in 1966. He has a PhD in Mathematics from the Universidad Complutense de Madrid (UCM) in 2002. He is currently a lecturer in the Department of Artificial Intelligence of the Universidad Nacional de Educación a Distancia (UNED), Spain, where he joined in 2009. Previously, he was lecturer in the Department of Statistics in the Universidad Carlos III de Madrid (UC3M) during 5 years. He has also worked for 15 years in the Department of Planning and Statistics in the banking sector and the pharmaceutical industry. His research interests include Survival Analysis, Nonparametric Tests, PLS, Meta-Analysis, Biostatistics and B-Learning. In @emilioleton, his professional tweets and his personal website with extended information can be found.

**Félix Hernández-del-Olmo** received the BSc degree in Physics from the Universidad Autónoma de Madrid, Madrid, Spain, in 1996 and Engineering and PhD degrees both in computer science from the National Distance Universty of Spain (UNED), Madrid, Spain, in 1999 and 2005, respectively. He is currently an Associate Professor with the Department of Artificial Intelligence, UNED. He has authored or coauthored more than 30 papers about educational systems, adaptive, and recommender systems in international journals and conferences. His research interests include educational systems, recommender systems, adaptive systems, and machine learning.

**How to cite this article:** Anaya AR, Luque M, Letón E, Hernández-del-Olmo F. Automatic assignment of reviewers in an online peer assessment task based on social interactions. *Expert Systems*. 2019;36:e12405. https://doi.org/10.1111/exsy.12405