

# Chapter 10

## Revealing Latent Student Traits in Distance Learning Through SNA and PCA



Rozita Tsoni, Evangelos Sakkopoulos, and Vassilios S. Verykios

**Abstract** Distance Learning has moved almost completely online, gaining ground in an educational setting of constantly increasing demand. Physical distance poses barriers in the implementation of such a transition, however, most of these barriers can be surpassed by implementing a Learning Analytics process around the educational process. The chapter presents a novel approach that is based on a rich spectrum of metrics of Social Network Analysis that can capture complicated interaction of social students' behavior, along with academic performance indicators, in a process that aims to reveal the latent characteristics of students participating in the discussion fora of their Distance Learning postgraduate course.

**Keywords** Higher education · Distance learning · Learning analytics · Social network analysis · Principal component analysis

### 10.1 Introduction

In the past decade, a large score of the learning and teaching activities have been transferred online. Recent technological and socio-economic developments on top of unpredictable global events pose even more imperatively the need for Open and Distance Learning. Restrictions for preventing Covid-19 infection led more than 1.5 billion enrolled students from all over the world (approximately 90% of the global student population) to experience interruption of education [52]. A massive, urgent transition of the conventional teaching and learning on the web increased the need

---

R. Tsoni (✉) · V. S. Verykios  
School of Science and Technology, Hellenic Open University, Patras, Greece  
e-mail: [rozita.tsoni@ac.eap.gr](mailto:rozita.tsoni@ac.eap.gr)

V. S. Verykios  
e-mail: [verykios@eap.gr](mailto:verykios@eap.gr)

E. Sakkopoulos  
School of Information and Communication Technologies, Department of Informatics,  
University of Piraeus, Piraeus, Greece  
e-mail: [sakkopul@unipi.gr](mailto:sakkopul@unipi.gr)

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022  
M. Ivanović et al. (eds.), *Handbook on Intelligent Techniques in the Educational Process*,  
Learning and Analytics in Intelligent Systems 29,  
[https://doi.org/10.1007/978-3-031-04662-9\\_10](https://doi.org/10.1007/978-3-031-04662-9_10)

for monitoring students' online behavior. Therefore, Learning Analytics (LA) was brought into the spotlight as the most promising tool to diminish the spatial and temporal distance between learning stakeholders.

Even without the massive disruption of Covid-19, there is an obvious upcoming change in the Higher Education setting. New models of teaching and learning have moved conventional systems from being able to service a small number of participants to massive, open courses, replacing a part of tutor's assistance and evaluation with automated or peer assistance and evaluation, resulting in a fewer percentage of students completing the courses successfully [28]. Higher education has emerged toward increased specialization and individualized instruction, while attention shifts from institutions and programs to individual students who aim to construct skill sets according to the new demands of the job market [15].

As a result, the complexity of the field is growing since personalization comes along with massive demand and large heterogeneity. Numerous students' communities are formed in large-scale courses offered by top-rated universities. Moreover, learning cannot be seen apart from social interaction whether this is happening implicitly or explicitly. Therefore, having a clear picture of tutors' and learners' interaction is vital to maintain and improve the quality of Distance Education. While teaching does not necessarily lead to learning [29], there is a constant need for feedback to evaluate the engagement of the learners and the effectiveness of the learning process [38]. In Distance Education tutors in online classes, without adequate information, may be misled by an unnoticed mismatch between ideal and actual class dynamics, from a social learning perspective [22]. Engagement and productive dialogue are important conditions for successful teaching. Network Analysis can identify where productive dialogue takes place [46]. Ferguson and Shum [17] found that educational success was correlated to the quality of learners' educational dialogue and students' satisfaction [37].

LA has the potential to: *“dramatically impact the existing models of education and to generate new insights into what works and what does not work in teaching and learning”* [41] Moreover, recently, the analysis of the existing evidence for LA indicates that there is a shift towards a deeper understanding of students' learning experiences [53]. LA focuses on the specific problem of understanding and optimizing factors that lead to a successful educational experience for all learners [31]. The implication of LA and the evaluation of its impact on learning is one of the key challenges of the educational field [45]. Also, one key principle in the *“Global guidelines for ethics in LA”* is to consider whether access to knowing and understanding more about how students learn brings with it a moral obligation to act [42].

There are many ways in which LA can impact education. A taxonomy of the LA types depending on their result is described by Downes [16] and contains descriptive analytics, diagnostic analytics, predictive analytics, prescriptive analytics, generative analytics, and deontic analytics. Additionally, in a research concerning the implementation of LA in countries identified seven major factors that should be taken into account: power, pedagogy, validity, regulation, complexity, ethics, and affect [18].

Our main stand in this context is to use LA for descriptive and diagnostic purposes mostly with a focus on identifying how the social behavior of students as it is

ascribed in their interactions with their fellow students and their tutors affects their learning. In our previous work [48, 51], qualitative conclusions were drawn by posing different research questions from well-established educational theories and using social network visualization and polarity analysis to answer them. In that respect, centrality measures and their distribution were studied in two-mode networks, and their projection onto one-mode networks, were extensively investigated [47].

In this chapter, we present a novel approach that is based on a rich spectrum of metrics of Social Network Analysis (SNA) that can capture complicated interaction of social students' behavior, along with academic performance variables, in a process that aims to reveal the latent characteristics of students participating in the discussion fora of their Distance Learning postgraduate course. Hopefully, actionable knowledge will be produced, helping tutors and educational stakeholders to base their decisions on the learners' needs even when these needs are not clearly stated.

This chapter is structured as follows: In Sect. 10.2 certain concepts and metrics mainly concerning Network Analysis are discussed. Additionally, the Hyperlink-Induced Topic Search (HITS) algorithm is presented along with a brief description of the Principal Component Analysis (PCA) and the Hierarchical Cluster Analysis techniques providing a complete report of the analytical tools used in the research. In Sect. 10.3 relevant work, concerning forum interaction and students' online behavior, is briefly presented. The following Sects. 10.4 and 10.5 describe the data and the experimental method used in our research. Results are presented and discussed in Sect. 10.6, providing educationally sound interpretation and understanding. In the final section conclusions, limitations, and future work are discussed.

## 10.2 Background: Definitions, Algorithms, and Methods

Our LA approach consists of two main steps. We can consider them as an analysis and a meta-analysis step. During the first step of our proposed methodology, we apply Network Analysis. Below we discuss relevant metrics, along with algorithms and methods that were used in this context. We also touch upon and introduce PCA and clustering, which are built around our primary analysis technique and used as a meta-analysis phase to build on the primary findings of SNA.

Network analysis includes concepts and metrics that allow the representation of the interaction between actors in the form of a network where each node represents an actor and the edge between two nodes represents connections (some kind of association). Here, we deal with networks containing either a single or a multiple type of nodes. When the nodes are of one kind the network is unimodal (or one-node network). Multimodal networks consist of different types of nodes. For example, forum participation can be expressed in two ways. In a one-mode network, where each node represents a participant, the edges indicate the interaction among them. In other words, a link indicates a comment or a reply to each other's post. Alternatively, the forum community can be imprinted as a two-mode network where nodes might represent participants or discussion threads. There, a link always connects two nodes

of a different type. A participant-node is connected directly only with thread-like nodes and vice versa. In the analysis that follows the representation of students' networks is bimodal as this allows for a more thorough view of the interaction, providing richer information than the projected one-mode network [47].

The reciprocity of the interaction among different nodes is associated with the feature of directivity. A network can be either directed or undirected. In a directed network, when a node  $i$ , is connected to a node  $j$ , the node  $j$  is not necessarily connected to node  $i$ . Thus, the adjacency matrix of the network is not necessarily symmetric. The links between a person and his/her followers in a social network site are an example of directed edges between the person and their followers because they are not reciprocal. Therefore, a network of Twitter accounts and followers is a directed network. On the contrary, an undirected network corresponds to a symmetric adjacency matrix. Thus, for every pair of nodes  $i, j$ ,  $a_{ij} = a_{ji}$ . A friendship network is undirected because of the mutuality of the relation.

Starting to explore a network  $G$ , the first and simplest metric describing a node is its *degree*. It is the sum of the number of edges connected to this node. In a bimodal network that represents a discussion forum activity, the high degree of a participant shows that s/he has posted in a lot of discussions but does not provide us any information about the number of the people he interacted with. The degree rises if the person posts in more discussions even if always the same person participates in them. In a directed network the *in-degree* is the number of incoming edges, and the *outdegree* is the number of outgoing edges. Forum interaction can be represented as a bimodal directed network where all edges come out of person-nodes and point to discussion-nodes. Therefore, the indegree of person-nodes and the outdegree of discussion-nodes are always zero.

The *Weighted degree* metric counts the number of edges but also adds up the number of times a person has posted in a certain discussion in a bimodal network.

*Closeness centrality* shows how many "hops" are needed so that a node can reach other nodes. The concept of geodesic distance is necessary to define closeness centrality. Geodesic distance is the number of edges that contain the shortest path between two given nodes. Closeness centrality is inversely proportional to the total geodesic distance from a node to all other nodes of a network [19]. Since closeness centrality depends on the networks' magnitude (the number of nodes and edges that contains), it is strongly affected by the network's type. In a bimodal network, for a given person  $A$  to reach another person  $B$  a discussion node has to intercede between them. Therefore, the distance between any two nodes cannot be less than two. The closeness centrality of a vertex is defined as:

$$C(x) = \frac{N - 1}{\sum_{x \neq y} d(x, y)} \quad (10.1)$$

where  $x, y$  are two vertices in the network  $G$ ,  $d(x, y)$  denotes their distance, and  $N$  is the number of nodes in the graph. Thus, a node with low closeness centrality is a

central node in the network. In other words, if the sum of the distances is large, then the closeness is small and vice versa [33].

*Harmonic closeness centrality* which is, likewise the Closeness centrality, a measure of nodes' proximity was proposed by Marchiori and Latora [32]. Harmonic closeness centrality can be computed in a not necessarily complete network because in case that there is no path between  $x, y$  the value of  $\frac{1}{d(y,x)}$  equals zero.

It is defined as:

$$H(x) = \sum_{y \neq x} \frac{1}{d(y, x)} \tag{10.2}$$

A nodes' *betweenness centrality* is a metric indicating the node it is contributing to connecting other nodes. This measure provides information about the "importance" of a node that despite the fact that it is a low degree node, it happens to be in a strategic location within the network. Nodes with high betweenness centrality act as bridges linking sub-groups of nodes that otherwise might be disconnected. In a discussion forum network, tutors have high betweenness centrality holding together the learning community [47].

*Betweenness centrality* of a given node  $i$  is proportional to the total number of geodesics between two given nodes  $j$  and  $k$ , which include node  $i$ . A node  $i$  with high betweenness centrality is part of many paths of its network. In general, when more ties are added in a network with a given number of nodes, betweenness centrality decreases [4]. Betweenness centrality can be expresses as:

$$BC(x) = \sum \frac{\sigma_{(v,w)}(x)}{\sigma_{(v,w)}} \tag{10.3}$$

where  $x$  denotes a given node,  $\sigma$  denotes the count operation,  $\sigma_{(v,w)}(x)$  is the number of shortest paths (between any pair of nodes  $v, w$  in the graph) that passes through the target node  $x$ , and  $\sigma_{(v,w)}$  the total number of shortest paths that occur between any pair of nodes of the graph. The target node would have a high betweenness centrality if it appears in many shortest paths [34].

*Eigenvector centrality* is a node's metric strongly affected by its neighbor's characteristics. It is defined as follows: given the adjacency matrix  $A$  of the network  $G$ , an eigenvector for this matrix is a vector  $v$  that satisfies the matrix–vector equation  $Av = av$  for some scalar value  $a$  (the eigenvalue). This would give the equation:

$$ECx_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j \tag{10.4}$$

where  $\lambda$  is a constant and  $j \in M(i)$  means that the sum is over all  $j$  such that the nodes  $i, j$  are connected (where  $M(i)$  denotes the set of all the nodes that are directly connected with node  $(i)$ ). Eigenvector centrality indicates a node's influence which

signifies its strategic position in a network. Highly influential nodes are connected with other nodes of high influence, adding value to each other.

*Eccentricity* is a distance measure that is considered to be much simpler than closeness centrality [39]. In a given network  $G$  the eccentricity  $e_{G(v)}$  of a node  $v$  is the maximum distance between node  $v$  and node  $u$  over all the nodes of the network. From the definition below (Eq. 10.5) it is obvious that a node with high eccentricity is a distant node.

$$e_{G(v)} = \max\{dist_G(v, u) : u \in V\} \quad (10.5)$$

A more advanced approach to investigate nodes' strategic position and influence is through a sophisticated approach, the so-called HITS algorithm. HITS, also known as *Hubs and Authorities* is a link analysis algorithm initially created to rank web pages [26] on the internet. The algorithm assigns two values in each node: a hub value and an authority value. A high value of hub means that the node points to high authorities i.e., nodes with valuable information. Respectively, a node with high authority is being pointed by good hubs in a mutually reinforcing relationship. The computation of these values is based on an iterative process that follows the principle of repeated improvement as a good hub adds value to an authority and subsequently, the authority adds more value to the hub in a repeated process that converges to a final result. The degree of convergence  $\epsilon$  (epsilon) determines the ending point of the iterative algorithm that is the maximum divergence between two sequential results. Initially, for each node  $p$  we set  $x^{<p>} = 1$  and  $y^{<p>} = 1$  for the ranking process to begin (where  $x^{<p>}$  denotes the hub and  $y^{<p>}$  the authority of the node. The function that updates the weights for hubs and authorities (hub and authority update rules respectively) are:

$$x^{<p>} \leftarrow \sum_{q:(q,p) \in E} y^{<p>} \quad (10.6)$$

and

$$y^{<p>} \leftarrow \sum_{q:(q,p) \in E} x^{<p>} \quad (10.7)$$

A directed edge  $(p, q)$  indicates the presence of a link from  $p$  to  $q$ . In a directed bimodal network of participant-to-thread interaction, an edge is always directed from a participant-node to a thread-node. Thus, participant-nodes have non-zero outdegree and always zero indegree. Additionally, thread-nodes always have zero outdegree and non-zero indegree. Therefore, participants are hubs with zero authority and threads are authorities with zero hubs.

The metrics that were briefly described above imprint different properties of each node in a network. Students' interaction and behavior within an online learning community is a multidimensional problem. The network metrics shed light into

the social aspect of learning. The online presence and the academic performance are revealed by the number of views and the grades of the students. There are many other features that affect the learning process such as personal and cultural information. However, the scope of the proposed methodology is to provide an accurate description of the learners based on the available data, preserving students' privacy and, at the same time, capturing a large range of their characteristics. To overcome the problems posed by the multidimensionality that was created from the incorporation of network metrics in our analysis it is useful to apply a method for dimensionality reduction. Exploratory Factor Analysis is a method that can prevent biased and skewed results, which are difficult to interpret and, additionally, can reveal hidden aspects of this multidimensional problem. Thus, most relevant metrics and variables were used in a Principal Component Analysis. PCA provides variables reduction, maintaining the majority of information. As an orthogonal linear transformation that projects the data to a new coordinate space, it produces main components that can be expressed as a linear combination of the initial variables weighted by their variance. The components that emerge are represented in a new orthogonal dimension revealing patterns and latent characteristics, that were not obvious in the first level of the analysis. Additionally, we can create graphs using the eigenvectors as new uncorrelated variables.

To further leverage the new normalized variables developed by PCA's factor scores, a clustering process is proposed. Hierarchical Cluster Analysis can provide additional information about students' behavior based on their latent characteristics. Thus, the method of between-group linkage using the Euclidean distance will be used in the new orthogonal space that PCA produced.

### 10.3 Related Work

Forum participation and interaction is a field of interest in education research because it reflects the relationships in the learning community and it can indicate issues where action must be undertaken by the learning facilitators. In Distance Learning it is crucial for the students to be and feel supported and, at the same time, to increase their autonomy. Descriptive and predictive models can help tutors to focus their attention on students at risk and prevent poor learning results. The activity of the students in the discussion fora was used for academic performance prediction by Chiu and Hew [10] along with views and posts count. Their research demonstrated greater predictive power in views count than in posts counts. Crossley et al. [11] shown that students had significantly better achievement than their peers when they made at least one post of 50 words or more. Furthermore, students who produce more on-topic posts, posts that are more strongly related to other posts, or posts that are more central to conversation presented a better completion rate. Sun et al. [44] compared the forum interaction between who participated in pre-defined groups and students in self-selected groups. It was found that there is a significant difference between the strength of the ties that students formed that led them to the conclusion

that the course design approach is affecting students' community structure. Chiru et al. [9] proposed a model for counting the strength of students' connection with certain discussion topics and with other participants, called the participant-topic and the topic-topic attraction.

In the field of Social Network Analysis visualization and metrics contribute to providing a deeper sight into the community structure, revealing relations and participants with a strategic role. Network Analysis of forum activity was considered in two successive studies in Hellenic Open University (HOU) [24, 30] aiming to create students' profiles based on their online participation in order to provide useful feedback to their tutors. Network representation through time reveals the evolution of students' community and along with polarity analysis can provide insights into the social aspect of their learning behavior [49]. In a literature review by Cela et al. [7] the most common metrics were found to be centrality and density, leading to the conclusion that Social Network Analysis, particularly when combined with content analysis, can provide a detailed understanding of the type of interactions between the members of the network, allowing the optimization of the course design. Hernández-García et al. [22] highlighted the need for tailored tools for advanced and in-depth analysis that will allow the effective confrontation of the problems that commonly appear in Distance Learning.

De-Marcos et al. [14] used network metrics to conduct PCA and also, they examined the correlation between those metrics and academic achievement. An analysis of the data retrieved from a social networking site that was delivered to students providing gamified activities and enabling social interaction and collaboration, showed a moderate correlation between most centrality measures and learning achievement. Thus, they concluded that structural metrics can be used as predictors of the learning outcome.

Tools and applications have been developed focusing on social network analysis and community detection [2, 3, 5, 8, 12] with different features [43] that can be used in educational research.

SNA was found to be more revealing concerning students' interactions and their location in the learning community [54]. This result is consistent with the research of Traxler et al. [46] that found that centrality measures were more reliable indicators of the grade than non-network measures such as post count. In a pedagogy-oriented work, Jan and Vlachopoulos [23] explored social network features as indicators of the structure of communities of practice and communities of inquiry. Their study substantiates the proposed Integrated Methodological Framework based on SNA as an effective framework for structural identification of community-based learning. A collaborative forum-based learning design, including pre-learning and post-learning activities, was proposed by Amano et al. [1]. In order to discover relevant structures in social networks generated from student communications Rabbany et al. [36] introduced a toolbox which automatically discovers relevant network structures, visualizes overall snapshots of interactions between the participants in the discussion forums, and outlines the leader and peripheral students.



While LA can be applied in a broad range of educational fields providing answers to a variety of students' related problems [25] research questions that contain structured hypotheses are not always the case. The computational, data-driven research is gaining ground, especially when Machine Learning techniques are used. In the so-called "black-box" approach data lead to the hypothesis creation, revealing hidden aspects of the problem. On the contrary, the "white-box" is a theory-driven, hypothesis-testing statistical analysis. Taking into account the multidisciplinary and the special characteristics of the educational research Sharma et al. [40] proposed a combined method, the grey-box approach where data collection, feature extraction, feature selection, prediction, and interpretation form a pipeline that can be fine-tuned with specific research needs. Pedagogical criteria lead Gkontzis et al. [21] to divide the academic year of students into six periods before applying Machine Learning techniques for academic performance prediction.

Prabhakar and Zaiane [35] presented a framework using a hybrid particle swarm optimization to form student groups based certain attributes (like age, grade etc.). They argue the algorithm that they proposed can cluster students into dynamic learning groups and could be used for automated grouping in MOOCs. PCA was used to investigate the relation between the use of web2.0 tools and students' performance [20]. Certain activities were identified as indicators of students' success even though no significant correlations between learning styles and performance were found.

## 10.4 The Hellenic Open University Dataset

The HOU is the only Higher Education Institute that offers full Distance Learning courses at undergraduate and postgraduate level in Greece. Moreover, it is the only university in the country that admission does not require written exams. Students that meet some basic requirements can enroll, even if they live in remote geographical areas, as they do not have any obligation of physical presence, except for some laboratory courses held during summer vacation time in certain programs. These two characteristics are considered to be the main assets of the openness culture that this university represents. However, all study programs offer the opportunity of face-to-face meetings, mainly for advisory and motivational purposes. Communication between students and their tutor can be achieved via synchronous teleconference meetings and telephone calls, SMS text messages, e-mails, or through the discussion forum of the course. Peers officially interact only through the discussion forum although they usually create groups in social media to communicate outside of the formal learning environment. Forum interaction takes place with no external motivation. Students do not gain extra credits and forum-related assignments are not usually assigned. They post online only if they feel that they want to communicate with their peers for any reason. Thus, discussion forum topics may vary from general questions or statements to specific course-related questions.

Graduate programs at the School of Science and Technology, are usually consisted of courses that last an academic year. Each academic year students can attend up to five OSS (Group meeting for consulting) and have to hand over up to six written assignments in each course. These assignments are obligatory and the average grade (it has to be above 5/10) determines whether a student would be permitted to sit on the final exams or not. In some courses, there are quizzes and online tests available that can contribute a percentage of the final grade, but this feature usually varies.

A compulsory course offered in the first year of studies was chosen for the experimental evaluation of the proposed methodology because it represents the beginning of students' learning experience and their first contact with Distance Learning for a large number of them. In the academic year, 2019–2020 students were divided into seven groups. Each group had a different tutor-consultant in charge. Two of the groups participated in face-to-face meetings held in the two bigger cities of Greece (Athens and Thessaloniki) and the five other groups had their synchronous meetings online (we will refer to them as e-groups). Students had to choose whether to participate in online or in face-to-face meetings at the beginning of the academic year for the groups to be formed. Data were drawn from the forum that was accessible from students and tutors of all groups, hosted in Moodle platform at the School of Science and Technology of the HOU in the academic year 2019–2020. Assignments' grades, final grades, views, and forum data were included in the dataset used for the analysis.

A high priority is the data mining and data management process to comply with the newly established General Data Protection Regulation. Privacy protection applies in all stages of data processing: data preprocessing, data analysis, and data publishing [27]. Thus, data went through the process of anonymization and a reference number was assigned to each student.

## 10.5 Data Analysis Process

The preliminary steps of the analysis included descriptive statistics and the creation of new data tables with binary variables (e.g., pass/fail, participant/non-participant). This is the stage of the formation of a detailed description of the participants, attaching additional features to the nodes, in order to follow the analysis of the interaction network.

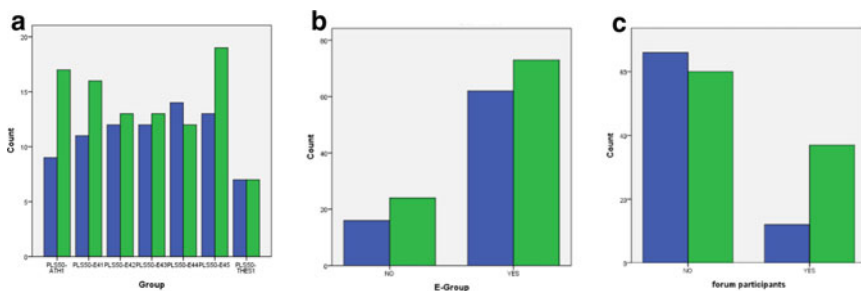
There is a threefold benefit of using Social Network Analysis in educational data. Firstly, the visualization of the network, that allows the investigation of the relations in the forum community, the type and the strength of the bonds between participants, and the positions of its members. Secondly, metrics about each participant and the complete network accurately determine the features of the participants and the structure of their community. Different roles and attributes that constitute meaningful information about the participants' behavior can be identified. Thirdly, metrics that were derived from SNA provide rich information and can be used for further analysis (Fig. 10.1).

The main characteristics of the students' community that emerged from the descriptive analysis are presented in Table 10.1. Hereupon, the overall academic performance of the students was seen in the context of different features that possibly affect it. The final grade is considered to be the most typical metric of students' performance. Students of HOU in order to pass the course have to have a minimum score of 50% in the final exams. Additionally, in case of failure or absence, they have a second chance in the re-examination. Therefore, a cross-tabulation was conducted between students' final achievement (the binary variable "pass or fail") and three variables ("meeting group" depending on which of the seven groups a student belongs, "e-group" which is the binary grouped variable of the groups that have online meetings and the groups that have face to face meetings, and "forum participant" that is also a binary variable depending on whether a student participated in the forum or not). Results revealed a statistically significant difference ( $p < 0.01$ ) in students' success for the students that participated in the forum, while the results did not vary significantly within different groups or whether students were participating in the online meetings or the face-to-face meetings.

Therefore, data were modified to fit Gephi's requirements into nodes' and edges' tables and were loaded in the application. Force Atlas algorithm was used to achieve

**Table 10.1** Basic information about the students

Number of students	175
Number of academic staff	10
Forum participants (students)	49
Number of students in the online meeting groups	135
Number of students in the f2f meeting groups	40
Forum participation rate	28%
Number of posts	1309
Number of threads	179
Average number of posts per participant	22.2

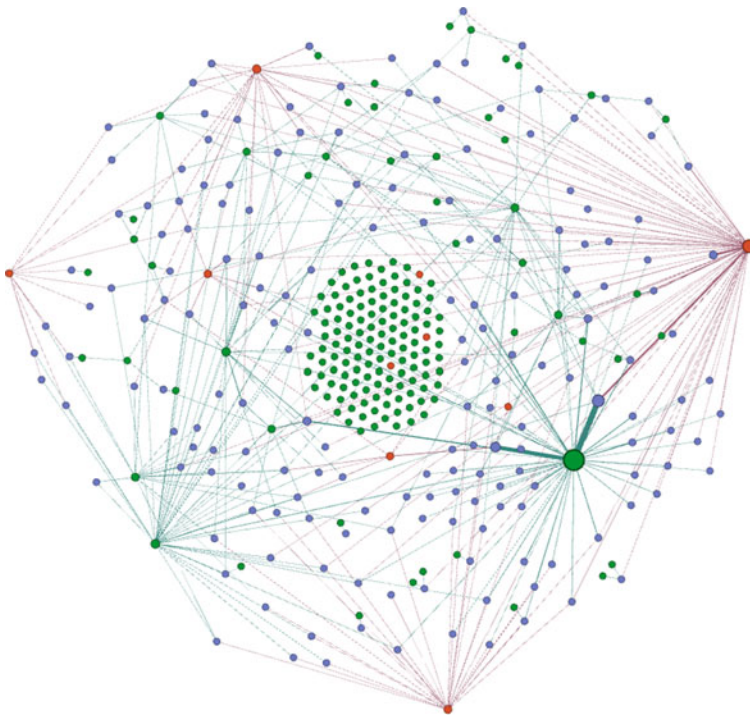


**Fig. 10.1** **a** Students' achievement per meeting group, **b** Students' achievement: online and face to face groups, **c** Students' achievement: forum participants and participants (green: pass, blue: fail)

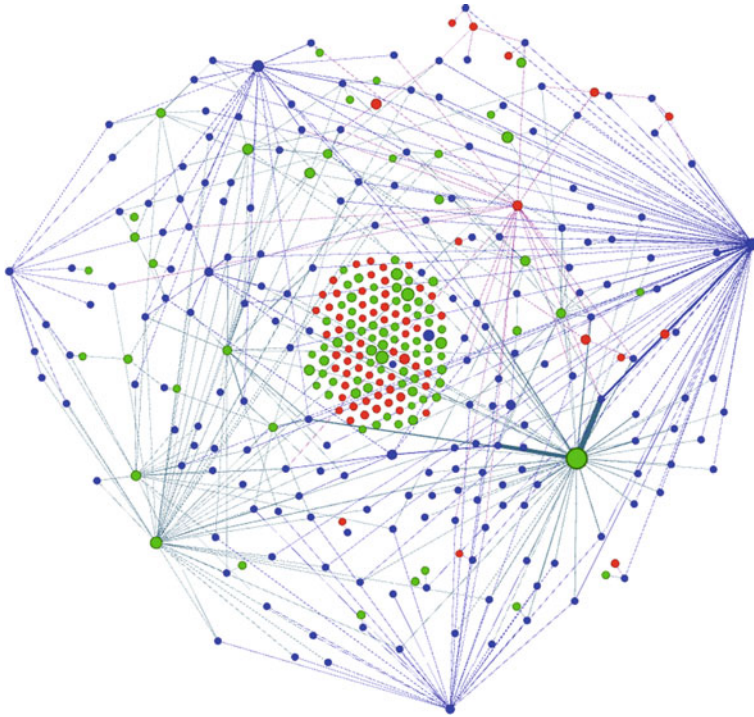
a readable visualization with no overlapping nodes and a minimum number crossing edges based on gravity, repulsion, and inertia [2]. Subsequently, certain partition and ranking methods were applied to assign different colors and magnitude to nodes and edges according to chosen variables and allowing multiple visualizations of the same network, highlighting different features each time (Figs. 10.2 and 10.3).

Fifteen variables were chosen to be included in the PCA that constitutes the next step of the analysis. The grades of the 6 written assignments (variables: WA\_1, WA\_2, WA\_3, WA\_4, WA\_5, WA\_6), the final grade, the number of views, and seven more network metrics (degree, weighted degree, betweenness centrality, harmonic closeness centrality, eigenvector centrality, hub, and eccentricity). Although in the first steps of the analysis all of the participants were included, for PCA the academic staff was excluded for two main reasons:

- (a) Students and tutors belong to different groups of participants, with different features and behavior. The scope of this research is to investigate students' actions and reveal latent characteristics that hopefully will be used to improve the educational process. The investigation of tutors' behavior demands a different methodological approach.
- (b) Seven of the variables chosen for the analysis do not apply in the case of tutors.



**Fig. 10.2** The participants' network annotated by the node's type and weighted degree (purple nodes: threads, green nodes: students, orange nodes: tutors)



**Fig. 10.3** The participant's network annotated by final exam's result and views count

As a last step of the proposed methodology, the factor scores that emerged from the PCA were saved as variables and were used to conduct hierarchical cluster analysis and graphical representation. Briefly, the methodological steps that describe our experimental evaluation are technically the following:

- i. Retrieve log files from forum activity in Moodle Platform
- ii. Anonymize data
- iii. Clear data and keep only every event's id. Each student is represented by a unique id number
- iv. Create edges files and nodes files with nodes annotation
- v. Import data to Gephi
- vi. Run Force Atlas algorithm for network formation
- vii. Run statistical measures for the bimodal network
- viii. Use partition to format the network
- ix. Export results containing network measures
- x. Use the new dataset for further analysis: descriptive statistics, cross-tabulations, distributions, correlations, clustering, and principal components analysis

## 10.6 Explicit and Latent Characteristics of the Students' Community

The results can be grouped into four levels deriving from the methodology, presented above:

- a. Descriptives and Cross tabulations
- b. SNA with Gephi software and metrics distribution
- c. Correlations
- d. PCA and clustering

### 10.6.1 *The Descriptive Features of Students' Community*

The main characteristics of the students' community that emerged from the descriptive analysis are presented in Table 10.1. Hereupon, the overall academic performance of the students was seen in the context of different features that possibly affect it. The final grade is considered to be the most typical metric of students' performance. Students of HOU in order to pass the course have to have a minimum score of 50% in the final exams. Additionally, in case of failure or absence, they have a second chance in the re-examination. Therefore, a cross-tabulation was conducted between students' final achievement (the binary variable "pass or fail") and three variables ("meeting group" depending on which of the seven groups a student belongs, "e-group" which is the binary grouped variable of the groups that have online meetings and the groups that have face to face meetings, and "forum participant" that is also a binary variable depending on whether a student participated in the forum or not). Results revealed a statistically significant difference ( $p < 0.01$ ) in students' success for the students that participated in the forum, while the results did not vary significantly within different groups or whether students were participating in the online meetings or the face-to-face meetings.

### 10.6.2 *Social Network Analysis and Distributions*

The second level of analysis provided information-rich graphs along with important metrics about the social aspect of learning. The first network's graph (Fig. 10.2) presents the interaction network that contains two types of nodes. Participant-nodes and threads-nodes (purple dots). Participants can either be students (green dots) or tutors (orange dots). There is also a ranking in the nodes' magnitude by their weighted degree (min 10-max 30) that indicated their overall participation. It is obvious that some central nodes gather around them lower degree nodes. Also, there is a very distinct group of disconnected nodes in the center of the network. These are the students who don't participate in the forum community.

In Fig. 10.3 a different partition has been implemented. Green nodes represent students that passed the course, orange nodes represent students who failed and blue nodes represent tutors of discussion threads. Nodes' magnitude depends on views ranking (min 10-max 30). The majority of active students passed the course. On the other hand, some of the inactive students have high views count. That means that although they don't participate in the forum community by posting or replying to the discussion, they read the posts of their peers.

Network metrics provide a more accurate image of the interaction. In Table 10.2 the characteristics of the ten more active forum participants (by descending value of weighted degree) are presented. The most active participant is a student with a significantly higher number of posts than all the other participants. It is worth noticing that even though Std\_18 has posted far more messages and had much more views than the second more active participant (Tutor\_1), two important network metrics tell a different story. The betweenness centrality of Tutor\_1, which indicates the mediative role of a node, is higher. Also, the eigenvector centrality of Tutor\_1 is greater than the eigenvector centrality of Std\_18, meaning that Tutor\_1 has a more strategic position in the network, therefore, is a highly influential node.

Participants' Degree (Fig. 10.4) and Weighted Degree follow a power-law distribution. Additionally, most participants have a relatively low Betweenness centrality, indicating that there are a few nodes (Tutor\_1, Std\_18, Std\_51) in a mediative role in students' collaboration community (see Table 10.1).

Most of the participants had up to 1000 views during the academic year. However, there was a small number of participants with higher views number. It is worth noticing that a specific student had over 5000 views at the same time where the average number of his/her peers, including the academic staff, was approximately 430 views. The larger number of participants with low views count is consistent with the relatively low forum participation rate (Table 10.1). The visualization of

**Table 10.2** Characteristics of top ten more active forum users

Participant	Degree	Weighted degree	Harmonic closeness centrality	Eigenvector centrality	Betweenness centrality	Views	Final grade
Std_18	57	483	0,49	0,76	6.939,39	5139	8
Tutor_1	68	209	0,55	0,92	11.505,06	1652	N/A
Std_51	42	68	0,46	0,55	5.176,13	1771	8
Std_46	21	63	0,39	0,25	1.604,65	791	7,7
Tutor_5	23	56	0,40	0,25	2.679,92	1510	N/A
Std_43	21	41	0,39	0,23	2.152,22	1006	0
Tutor_6	28	40	0,39	0,27	2.421,97	605	N/A
Tutor_9	10	30	0,35	0,13	461,89	602	N/A
Std_9	7	27	0,33	0,07	731,39	629	6,7
Tutor_7	8	26	0,32	0,05	569,19	956	N/A

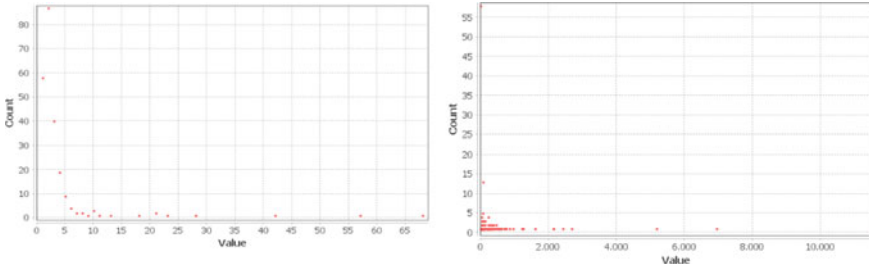


Fig. 10.4 Degree Distribution (Left), Betweenness centrality Distribution

the views counts also follows a Power Law distribution. The distribution of the final grade was also visualized (Fig. 10.5).

The final grade is the highest of the grades of the final exams and re-examination. 59 students are appearing to have a final grade equal to zero. This number represents the students who did not show up in any of the two examinations. The majority of them (61%) did not hand over written assignments 4, 5, and 6, proving that they practically dropped out of the course earlier, in the mid academic year. Additionally, some of the students fail to achieve an average grade of 5/10 in the 6 written assignments resulting in their exclusion from the final examination.

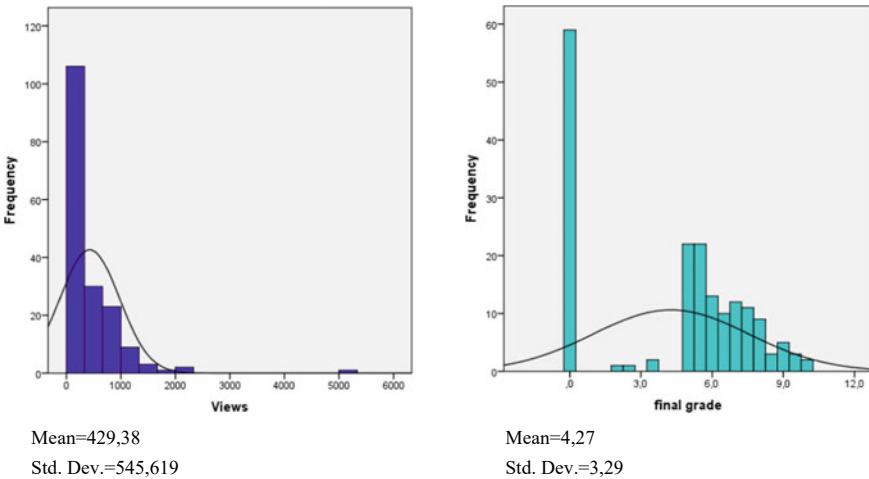


Fig. 10.5 The distribution of views count (left), and the distribution of students' final grade (right)



### 10.6.3 Correlations

In the next level of the analysis, the correlation of students' grades with variables that describe their online activity and their position in the forum community revealed some interesting results. There was a strong positive correlation between the grades of the 2nd WA ( $r = 0,597$ ,  $p < 0,05$ ) and the final grade and a moderate positive correlation between the grades of the 1st ( $r = 0,496$ ,  $p < 0,05$ ), 3rd ( $r = 0,451$ ,  $p < 0,05$ ), 4th, ( $r = 0,466$ ,  $p < 0,05$ ), and 5th WA ( $r = 0,487$ ,  $p < 0,05$ ), and the final grade. This result signifies that the 2nd WA is the most representative of the final achievement. There was no significant correlation between the grade of the 6th WA and the final grade ( $p > 0,05$ ). This, seemingly contradictory, result can be justified by the course's restrictions on participation in final exams. Many students who have ensured the right of participating in the final exam (that is an average grade of 5/10 in the WA) already in the 5th WA, minimize their effort in the 6th WA.

The views number also is moderately correlated to the final grade of the students ( $r = 0,298$ ,  $p < 0,05$ ). The position of the participants in the communication network does not seem to affect their final grades. It has to be noted that there were no structured learning activities given to the students that included forum interaction. In this sense, forum interaction is not directed linked to their academic performance. The instructional design significantly affects the expected results of SNA. In other words, a discussion forum where students have to complete certain learning tasks is expected to reflect the academic characteristics of the participants. However, in this case, where the discussion forum that is used for communication not restricted by topic, and does not yield grading rewards, the network structure reveals social characteristics rather than academic.

Although the final grade isn't related to network measures, the third (WA\_3) and the sixth written (WA\_6) assignment are. The WA\_3 is moderate negatively correlated ( $r = -0,360$ ,  $p < 0,05$ ) to eccentricity. Thus, most peripheral nodes had lower grades in the third written assignment than their most central peers. Interestingly, the sixth written assignment, although does not reflect the final achievement, is positively correlated with two important indices of the network: eigenvector centrality ( $r = 0,287$ ,  $p < 0,05$ ) and betweenness centrality ( $r = 0,302$ ,  $p < 0,05$ ). As was mentioned above, the WA\_6 differs from the previous assignments because of the exam's participation restriction. Therefore, two polar groups of students were formed. The high-graded students who already established the minimum 5/10 and chose to minimize their effort, and the low-graded students who are at risk of losing the right to participate in the exams, so they maximize their effort. As a result, they address to the discussion forum their questions, trying to get all the help that they need to raise their grades. High betweenness centrality indicates interaction with many different inner groups and high eigenvector centrality signifies communication with highly influence participants, where, in most cases, are tutors answering questions.

**Table 10.3** KMO and Bartlett’s Test

Kaiser–Meyer–Olkin Measure of Sampling Adequacy		0,754
Bartlett’s Test of Sphericity	Approx. Chi-Square	4190,699
	Df	105
	Sig	0,000

### 10.6.4 Factor Analysis and Clustering

Variables concerning students’ performance (WA\_1, WA\_2, WA\_3, WA\_4, WA\_5, WA\_6, and final grade) along with variables related to students’ online activity (views) and their position in the communication network (Degree, Weighted Degree, Eccentricity, Betweenness centrality, Harmonic closeness centrality, Hub and Eigenvector centrality) were imported for Principal Component Analysis. Firstly, a Kaiser–Meyer–Olkin (KMO) and Bartlett’s test was conducted to investigate the adequacy of our dataset for factor analysis through PCA. KMO score (Table 10.3) can be considered as middling suitable for PCA. Bartlett’s test results (Table 10.3) indicate that factor analysis can be used in these variables, as there are interrelated to a significant level.

Three major components were found, explaining approximately 79% of the total variance of the sample (Table 10.4). In Fig. 10.6 the scree plot of the eigenvalues of each component is presented. Only three of the eigenvalues are greater than 1, the rest of them explain a very small proportion of the variance, less than each initial variable explained, hence they can be rejected.

In order to explain and interpret the results of PCA in the educational context the rotated component matrix is used (Table 10.5). Varimax with Kaiser Normalization was the rotation method that converged in four iterations. Each of the three components is most highly correlated with a different group of variables that reflects another aspect of students learning presence. Component 1, which explains 43,65% of the variance, is highly correlated with metrics that concern forum interaction (views, degree, weighted degree, betweenness centrality, hubs, and eigenvector centrality). This component is about the social status of the students in the learning community that is expressed through the discussion forum. It sums up indicative features of the influence, the collaboration, the extroversion, and the participation of peers. Component 2, which explains 25,73% of the variance, is highly correlated with all the grading variables (WA\_1, WA\_2, WA\_3, WA\_4, WA\_5, WA\_6, and Final

**Table 10.4** Total variance explained

Component	Initial eigenvalues		
	Total	% of Variance	Cumulative %
1	6,548	43,651	43,651
2	3,859	25,725	69,375
3	1,452	9,679	<b>79,054</b>

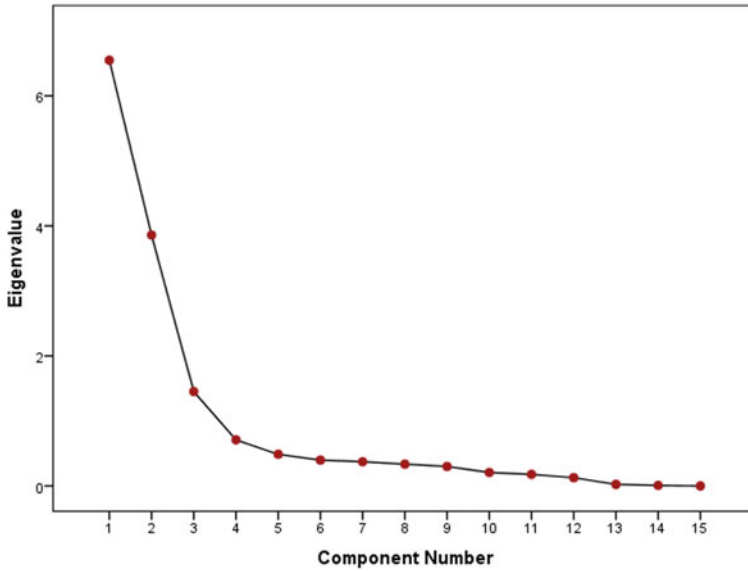
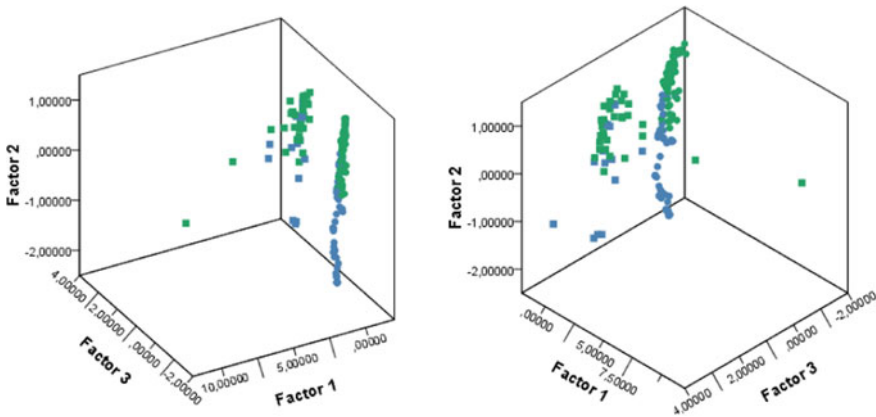


Fig. 10.6 Scree Plot

Table 10.5 Rotated component matrix

Variable	Component 1	Component 2	Component 3
WA_1	0,009	<b>0,698</b>	0,209
WA_2	0,030	<b>0,772</b>	0,144
WA_3	0,084	<b>0,798</b>	0,070
WA_4	0,086	<b>0,886</b>	0,003
WA_5	0,105	<b>0,847</b>	0,080
WA_6	0,165	<b>0,753</b>	-0,076
Final grade	0,087	<b>0,826</b>	0,046
Views	<b>0,734</b>	0,414	0,035
Degree	<b>0,950</b>	0,086	0,232
Weighted degree	<b>0,925</b>	0,031	0,032
Eccentricity	0,138	0,151	<b>0,923</b>
Betweenness centrality	<b>0,969</b>	0,060	0,106
Harmonic closeness centrality	0,345	0,109	<b>0,879</b>
Hub	<b>0,977</b>	0,062	0,160
Eigenvector centrality	<b>0,967</b>	0,072	0,188

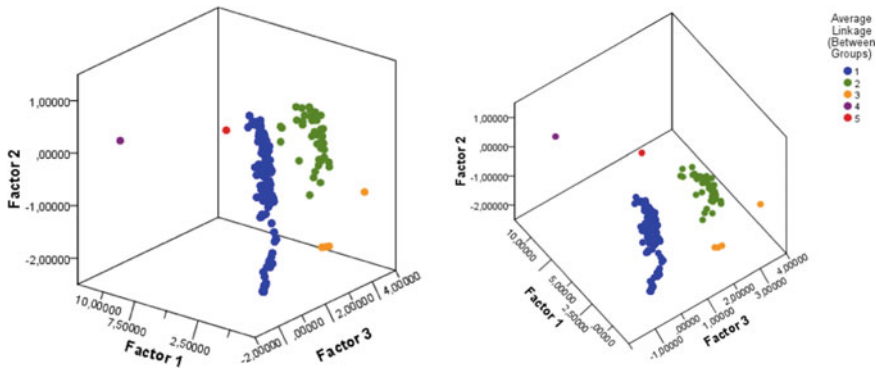


**Fig. 10.7** The 3D scatterplot of the three main components from two different angles (Partition: Green: Pass, Blue: Fail, Square: Forum Participant, Round: Non-participant)

grade). Thus, it reflects the academic profile of the students. Finally, component 3 that explains 9,68% of the variance, is highly related to two network measures: eccentricity and harmonic closeness centrality. These measures denote peripheral nodes in the network, therefore component 3 indicative for students who participate but tend to stay aside, in a loose engagement with the discussion community.

The visualization of students' eigenvectors can provide further explainable information. In Fig. 10.7, two different views of the three-dimensional graph are presented. Nodes are parted by color and by shape. The color defines whether the student passed (green nodes) or failed (blue nodes) the course. The shape separates forum participants (square nodes) from non-participants (round nodes). As it is shown in both graphs, the two main clusters corresponding to the two different groups of students. Those that participate in the forum community and collaborate online and those who see the educational process as a personal experience and do not have an online social presence. This result highlights the online social presence as the most important difference within students' behavior.

The clusters that emerged from the Hierarchical Cluster Analysis (Fig. 10.8) denoted two main groups that concentrate the majority of the students and three other clusters with outliers who present some notable features. Hence, cluster 1 contains the students with a low online presence (non-participants in the forum and low views count). Cluster 2 contains students who participated actively in the forum community. Cluster 3 contains students who followed a common path: initially participated and handed over the first one or two written assignments gaining high grades, however, at a later point, they dropped out of the course. Clusters 4 and 5 contain one node. Their profile explains their position in a different group as it is about the two most active students in the forum community. Both of them have excellent grades and over 1500 views. Additionally, they hold a central position in the network with high values in nearly all centrality measures. However, they are placed in different groups therefore



**Fig. 10.8** The 3D scatterplot of the three main components from two different angles colored by clustering group

there has to be a non-negligible difference between them. Their main difference relies on the number of views (5139 for cluster 4 and 1771 for cluster 5) and in their weighted degree especially compared to their degree (57 and 483 respectively, for cluster 4 and 42 and 68 respectively, for cluster 5). The difference between these numbers in a bimodal participant-thread network is the following: The first student (cluster 4) posted in 57 threads 483 times. That means that he/she posted approximately 8,5 times in each thread, while the second student (cluster 5) posted approximately 1,6 messages per thread. That explains why cluster 5 is located near the clusters of his/her peers and cluster 4 is relatively isolated.

## 10.7 Conclusions

The abrupt shift of many conventional institutions that moved their courses online without a thorough educational design, led Distance Education experts to make the distinction between Distance Education and “what is being practiced during the interruption of education, which can better be described as emergency remote education” [6]. A new challenge is therefore created for Distance Education institutions to lead the way towards “real” Distance Education rather than urged online courses, through quality students’ support. Thus, in an intriguing period for Higher Education, where Distance Learning is coping to establish its effectiveness, LA can provide answers to a series of aspects of teaching and learning such as students’ support, instructional design, policymaking, and educational leadership. While technology is increasingly providing tools the appropriate methodology to incorporate them in educational research and implementation is needed. This chapter describes a methodology that was implemented in data from a postgraduate distance learning course for revealing latent students’ characteristics, aiming to improve the educational process. Data from the online platform were analyzed and the network of students’ interaction

was created. It was shown that a combination of network metrics along with academic performance indices provided a detailed insight into the learning community and for individual students as well. Some of the results became obvious in the first steps of the analysis. However, PCA revealed three main factors: the “social status” factor, the “academic performance” factor, and the “loose engagement” factor.

The difficulties arising from the lack of physical presence in Distance Education can be addressed by effectively leveraging available data. Although several Learning Analytic approaches have been proposed, they usually focus separately on academic performance indices or in social network metrics. Additionally, SNA is used as the final step of the analysis, restricting its contribution in producing metrics to describe students’ community. Instead, we proposed a methodology that goes beyond simple description and uses SNA as a pre-processing step to retrieve metrics that would enrich our dataset preparing it for PCA and Clustering. This way, the clusters that emerge group students by latent traits, allowing tutors to focus on features that matter the most in the learning process. In our experiments, among others, cluster analysis identified a group of students in danger of dropping out who need further support (cluster 3), outliers that could disturb the collaborative spirit in the forum community (cluster 4), and also students who need external motivation to improve their effort levels (cluster 1).

As was mention above, educational research is a complicated, multi-dimensional field with various aspects that cannot all be captured by log file data. Therefore, we consider that the lack of personal, socio-economical, and cultural data along with data from previous learning experiences is the main limitation of our research. This is a difficult barrier to overcome, not only for practical but mainly for privacy reasons.

A student’s learning experience rarely starts and ends in one course. An integrative, complex, and holistic view is needed to understand the dynamics outside of a specific course that influence learning performance [13]. Therefore, we plan for the future to test this methodology in different datasets to compare results and discover patterns. Additionally, time series analysis and content analysis of the discussions should provide a deeper understanding of the interaction that takes place within the educational context.

## References

1. K. Amano, S. Tsuzuku, K. Suzuki, N. Hiraoka, Learning together for mastery by using a discussion forum, in *2019 International Symposium on Educational Technology (ISET)*, pp. 165–169. IEEE (2019)
2. M. Bastian, S. Heymann, M. Jacomy, Gephi: an open-source software for exploring and manipulating networks. *Icwsn* **8**(2009), 361–362 (2009)
3. V. Batagelj, A. Mrvar, Pajek-program for large network analysis. *Connections* **21**(2), 47–57 (1998)
4. S.P. Borgatti, M.G. Everett, Network analysis of 2-mode data. *Social networks* **19**(3), 243–270 (1997)
5. S.P. Borgatti, M.G. Everett, L.C. Freeman, *Ucinet for windows: software for social network analysis*. Harvard, MA: Analyt. Technol., 6 (2002)

6. A. Bozkurt, I. Jung, J. Xiao, V. Vladimirschi, R. Schuwer, G. Egorov, ... & V. Rodes, A global outlook to the interruption of education due to COVID-19 Pandemic: navigating in a time of uncertainty and crisis. *Asian J. Dist. Educ.* **15**(1), 1–126 (2020)
7. K.L. Cela, M.Á. Sicilia, S. Sánchez, Social network analysis in e-learning environments: a preliminary systematic review. *Educ. Psychol. Rev.* **27**(1), 219–246 (2015)
8. J. Chen, J. Fagnan, R. Goebel, R. Rabbany, F. Sangi, M. Takaffoli, ... & O. Zaiane, Meerkat: Community mining with dynamic social networks, in *2010 IEEE International Conference on Data Mining Workshops*, pp. 1377–1380. IEEE (2010)
9. C.G. Chiru, T. Rebedea, A. Erbaru, Using pagerank for detecting the attraction between participants and topics in a conversation, in *WEBIST* (1), pp. 294–301 (2014)
10. T.K. Chiu, T.K. Hew, Factors influencing peer learning and performance in MOOC asynchronous online discussion forum. *Australasian J. Educ. Technol.* **34**(4) (2018)
11. S. Crossley, M. Dascalu, D.S. McNamara, R. Baker, S. Trausan-Matu, *Predicting success in massive open online courses (MOOCs) using cohesion network analysis* (International Society of the Learning Sciences, Philadelphia, PA, 2017)
12. M.G. Csardi, Package ‘igraph.’ Last accessed **3**(09), 2013 (2013)
13. S. Dawson, S. Joksimovic, O. Poquet, G. Siemens, Increasing the impact of learning analytics, in *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pp. 446–455 (2019)
14. L. De-Marcos, E. García-López, A. García-Cabot, J.A. Medina-Merodio, A. Domínguez, J.J. Martínez-Herráiz, T. Diez-Folledo, Social network analysis of a gamified e-learning course: small-world phenomenon and network metrics as predictors of academic performance. *Comput. Hum. Behav.* **60**, 312–321 (2016)
15. R. DeMillo, New ecosystems for higher education and what they mean for accreditation and assessment. *WASC Concept Papers, Second Series: The Changing Ecology of Higher Education*. WASC (2013)
16. S. Downes, Ethical codes and learning analytics. *Hum. Artif. Intell. Soc. Future* (2020)
17. R. Ferguson, S.B. Shum, Social learning analytics: five approaches, in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, pp. 23–33. ACM (2012)
18. R. Ferguson, D. Clow, D. Griffiths, A. Brasher, Moving forward with learning analytics: expert views. *J. Learn. Analyt.* **6**(3), 43–59 (2019)
19. L.C. Freeman, A set of measures of centrality based on betweenness. *Sociometry*, pp. 35–41 (1977)
20. C. Giovannella, F. Scaccia, E. Popescu, A PCA study of student performance indicators in a Web 2.0-based learning environment, in *2013 IEEE 13th International Conference on Advanced Learning Technologies*, pp. 33–35. IEEE (2013)
21. A.F. Gkontziz, S. Kotsiantis, D. Kalles, C.T. Panagiotakopoulos, V.S. Verykios, Polarity, emotions and online activity of students and tutors as features in predicting grades. *Intell. Decis. Technol.*, 1–28 (2020)
22. Á. Hernández-García, I. Suárez-Navas, GraphFES: A web service and application for Moodle message board social graph extraction, in *Big Data and Learning Analytics in Higher Education*, pp. 167–194. Springer, Cham (2017)
23. S.K. Jan, P. Vlachopoulos, Social network analysis: a framework for identifying communities in higher education online learning. *Technol. Knowl. Learn.* **24**(4), 621–639 (2019)
24. V. Kagklis, A. Karatrantou, M. Tantoula, C.T. Panagiotakopoulos, V.S. Verykios, A learning analytics methodology for detecting sentiment in student fora: a case study in distance education. *Europ. J. Open Dist. E-Learn.* **18**(2), 74–94 (2015)
25. A. Klačnja-Milicevic, M. Ivanovic, Learning analytics-new flavor and benefits for educational environments. *Inform. Educ.* **17**(2), 285–300 (2018)
26. J.M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, A.S. Tomkins, The web as a graph: measurements, models, and methods, in *International Computing and Combinatorics Conference*, pp. 1–17. Springer, Berlin, Heidelberg (1999)
27. K.H. Kyritsi, V. Zorkadis, E.C. Stavropoulos, V.S. Verykios, Privacy issues in learning analytics. *Blended Online Learn.*, 218 (2018)

28. D. Laurillard, E. Kennedy, T. Wang, How could digital learning at scale address the issue of equity in education? Learning at scale for the global south. *Quezon City, Philippines: Foundation for Information Technology Education and Development* (2018)
29. A. Lionarakis, The theory of distance education and its complexity. *Europ. J. Open Dist. E-Learn.* **11**(1) (2008)
30. E. Lotsari, V.S. Verykios, C. Panagiotakopoulos, D. Kalles, A learning analytics methodology for student profiling, in *Hellenic Conference on Artificial Intelligence*, pp. 300–312. Springer, Cham (2014)
31. K. Mangaroska, M. Giannakos, Learning analytics for learning design: a systematic literature review of analytics-driven design to enhance learning. *IEEE Trans. Learn. Technol.* **12**(4), 516–534 (2018)
32. M. Marchiori, V. Latora, Harmony in the small-world. *Physica A* **285**(3–4), 539–546 (2000)
33. L. Metcalf, W. Casey, *Cybersecurity and applied mathematics*. Syngress (2016)
34. C. Perez, R. Germon, Graph creation and analysis for linking actors: application to social data, in *Automating Open Source Intelligence*, pp. 103–129. Syngress (2016)
35. S. Prabhakar, O.R. Zaiane, Learning group formation for massive open online courses (MOOCs). *International Association for Development of the Information Society* (2017)
36. R. Rabbany, M. Takaffoli, O.R. Zaiane, Social network analysis and mining to support the assessment of on-line student participation. *SIGKDD Explor.* **13**, 20–29 (2012)
37. Y.B. Rajabalee, M.I. Santally, Learner satisfaction, engagement and performances in an online module: Implications for institutional e-learning policy. *Educ. Inform. Technol.*, 1–34 (2020)
38. C. Reffay, T. Chanier, How social network analysis can help to measure cohesion in collaborative distance-learning, in *Designing for Change in Networked Learning Environments*, pp. 343–352. Springer, Dordrecht (2003)
39. J.S. Sereni, M. Krnc, R. Škrekovski, Z.B. Yilma, Eccentricity of networks with structural constraints. *Discussiones mathematicae*, 1–22 (2018)
40. K. Sharma, Z. Papamitsiou, M. Giannakos, Building pipelines for educational data using AI and multimodal analytics: a “grey-box” approach. *Br. J. Edu. Technol.* **50**(6), 3004–3031 (2019)
41. G. Siemens, Learning analytics: envisioning a research discipline and a domain of practice, in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, pp. 4–8 (2012)
42. S. Slade, A. Tait, *Global guidelines: ethics in learning analytics* (2019)
43. C. Sternitzke, A. Bartkowski, R. Schramm, Visualizing patent statistics by means of social network analysis tools. *World Patent Inf.* **30**(2), 115–131 (2008)
44. B. Sun, M. Wang, W. Guo, The influence of grouping/non-grouping strategies upon student interaction in online forum: a social network analysis, in *2018 International Symposium on Educational Technology (ISET)*, pp. 173–177. IEEE (2018)
45. J.P.L. Tan, E. Koh, Situating learning analytics pedagogically: towards an ecological lens. *Learn. Res. Pract.* **3**(1), 1–11 (2017). <https://doi.org/10.1080/23735082.2017.1305661>
46. A. Traxler, A. Gavrin, R. Lindell, Networks identify productive forum discussions. *Phys. Rev. Phys. Educ. Res.* **14**(2), 020107 (2018)
47. R. Tsoni, E. Sakkopoulos, C.T. Panagiotakopoulos, S.V. Verykios, On the equivalence between bimodal and unimodal students’ collaboration networks in distance learning. *J. Intell. Decis. Technol.* (2021)
48. R. Tsoni, V.S. Verykios, Looking for the “more knowledgeable other” through learning analytics, in *Proceeding of 10th International Conference in Open and Distance Learning* **10**(3A), 239–251 (2019)
49. R. Tsoni, E. Paxinou, E.C. Stavropoulos, C. Panagiotakopoulos, V. Verykios, Looking under the hood of students’ collaboration networks in distance learning. *The Envisioning Report for Empowering Universities* (2019)
50. R. Tsoni, C. Samaras, E. Paxinou, C. Panagiotakopoulos, V.S. Verykios, From analytics to cognition: expanding the reach of data in learning, in *Proceedings of CSEDU* (2019)
51. R. Tsoni, E.C. Stavropoulos, V. Verykios, Leveraging learning analytics with the power of words. *The Envisioning Report for Empowering Universities*, 24 (2019)



52. UNESCO, COVID-19 Educational disruption and response (2020). <https://en.unesco.org/covid19/educationresponse>
53. O. Viberg, M. Hatakka, O. Bälter, A. Mavroudi, The current landscape of learning analytics in higher education. *Comput. Hum. Behav.* **89**, 98–110 (2018)
54. N. Yusof, A.A. Rahman, Students' interactions in online asynchronous discussion forum: a social network analysis, in *2009 International Conference on Education Technology and Computer*, pp. 25–29. IEEE (2009)