Ben Kei Daniel *Editor*

Big Data and Learning Analytics in Higher Education

Current Theory and Practice

Springer

Big Data and Learning Analytics in Higher Education

angel.hernandez@upm.es

Ben Kei Daniel
Editor

Big Data and Learning Analytics in Higher Education

Current Theory and Practice

Springer

Editor
Ben Kei Daniel
University of Otago
Dunedin, New Zealand

ISBN 978-3-319-06519-9 ISBN 978-3-319-06520-5 (eBook)
DOI 10.1007/978-3-319-06520-5

Library of Congress Control Number: 2016947402

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Contents

Chapter 11
GraphFES: A Web Service and Application for Moodle Message Board Social Graph Extraction

Ángel Hernández-García and Ignacio Suárez-Navas

Abstract This chapter introduces GraphFES, a Web service and application that processes data from forum activity in Moodle courses and transforms them into social graphs to enable social learning analytics in Gephi, a social network analysis application. The chapter gives an overview of social learning analytics in online and computer-supported collaborative learning and describes existing tools for social network analysis of educational data. The chapter also presents the main concepts associated to the data source (Moodle logs) and target (Gephi), and a more detailed explanation of GraphFES's design and operation. An example with data from two courses illustrates how GraphFES and Gephi can combine to carry out social learning analytics in Moodle courses. The final section discusses the potential of this approach for effective social learning analytics.

Keywords Learning analytics • Social network analysis • Social learning analytics • Learning management systems • Forums • Computer-supported collaborative learning • Gephi • Educational data • Visualization • Moodle log

Introduction and General Context

The main difference between face-to-face—or even blended learning—and online-only instruction is the lack of physical interaction between teachers and students, and among learners. Apart from the results of assignments and eventual tutorial support sessions, in on-site (face-to-face) and mixed-method learning (blended learning) courses, instructors often rely on real-time feedback to get an idea of students' engagement and progress, at the individual, group, and course levels (Reffay & Chanier, 2003).

Á. Hernández-García (✉) • I. Suárez-Navas
Universidad Politécnica de Madrid, Madrid, Spain
e-mail: angel.hernandez@upm.es

B. Kei Daniel (ed.), *Big Data and Learning Analytics in Higher Education*,
DOI 10.1007/978-3-319-06520-5_11

However, the distinctive lack of physical interaction in online environments makes student tracking a difficult task for instructors and course coordinators. Without any means to analyze students' progress and online participation, there is a big risk that some students may fall behind without the teacher noticing it, and this situation can ultimately lead to student failure and attrition. Furthermore, lack of timely information may also lead to an unnoticed mismatch between ideal and actual class dynamics, from a social learning perspective.

This mismatch is especially important in computer-supported collaborative learning (CSCL) settings and/or courses with high student–teacher ratios. In these scenarios, monitoring student progress may be a difficult and time-consuming task for instructors due to the great amount of raw data available in current Virtual Learning Environments (VLEs) and Learning Management Systems (LMS) (Macfadyen & Dawson, 2012). In addition, emerging new instructional methods in which students also have social interactions as part of the learning process outside the formal learning contexts—e.g., personal learning environments (PLEs), social networking sites (SNSs)—or where the number of students is very large—e.g., massive online open courses (MOOCs)—make it necessary to provide teachers with tools to analyze the social dynamics of the class in online instruction.

The results of this analysis should give teachers enough useful and meaningful information at any given moment during the course to intervene, if necessary, or make fine-tuning adjustments to improve the whole learning process. The necessity of this type of analysis has led to the emergence of learning analytics as discipline. Learning analytics focuses on the collection, analysis, and reporting of educational data to better understand and optimize learning (Long & Siemens, 2011).

Hernández-García and Conde (2014) identify three main levels of learning analytics: identification of suitable indicators; identification, understanding, and explanation of learning behaviors; and mechanisms for adaptive learning. Although they are all related to each other, the second one is, by and large, the one that has raised most interest among scholars and practitioners—probably because it is the most immediate in terms of interpretation of results, and also has good value for theory building.

Most studies following this approach try to relate a student's activity in a technology-supported educational environment (e.g., LMS), many times under the assumption that "data speak for themselves," and neglecting other situational information, such as the assessment instruments employed or the social nature of the co-construction of knowledge in networks of practice (De Laat & Prinsen, 2014).

These omissions are even more relevant in courses where the instructional method relies heavily on collaborative and teamwork-based online learning because (1) it may be very difficult for instructors to detect dysfunctional groups or lack of student's engagement—especially in courses with a large number of enrolled students— which can ultimately lead to failure or course attrition, and (2) assessment in collaborative learning must include mastering of the instructional contents and participation (Barkley, Cross, & Major, 2005). Assessment in teamwork contexts is often just based on the result—the final work—as an evidence that can be measured and compared. Nevertheless, this assessment does not take into account individual participation of group members. In addition, gathering information about participation in online settings may lead to cumbersome analysis procedures (Fidalgo-Blanco, Sein-Echaluce, García-Peñalvo, & Conde, 2015).

In online collaborative learning, most of the data about students' participation is stored in the learning platform. Therefore, the application of social network analysis (SNA) to educational data—what Buckingham-Shum and Ferguson (2012) define as inherently social learning analytics—may offer insight on participation both from an analytical and visual standpoint.

This chapter introduces GraphFES (Graph Forum Extraction Service), a Web service for data extraction and processing of forum activity in a Moodle platform, and shows how data provided by GraphFES can be used for analysis and visualization of data about participation and engagement in Gephi—an open-source SNA software—(Bastian, Heymann, & Jacomy, 2009) in order to understand, explain, and improve learning processes in online contexts. As an example, the study includes the application of GraphFES to data from two different courses: a Master's course with few students and only a group assignment, and an undergraduate course based on a team project with high number of participants.

This chapter is structured as follows: Section "Social Learning Analytics" offers an overview of prior literature on social learning analytics. Section "Social Learning Analytics: Tools" presents different tools available for social learning analytics in online and ICT-supported learning. Sections "Moodle Logs and Data Extraction and Visualization," "Gephi: A Tool for Social Network Analysis," and "GraphFES: Design and Operation" detail Moodle's log data capabilities, the Gephi software, and the design and operation of GraphFES. Section "Case Studies" explains the characteristics of the courses used for the empirical study, the main results from the SNA and some visualizations of the resulting networks. Finally, section "Conclusion" will discuss the main findings of the study, addressing the limitations and future avenues of research on this topic.

Social Learning Analytics

Online learning systems give support to individual, self-directed learning by providing tools that enable access to learning resources, and by implementing assessment instruments and tools (quizzes, essays, etc.). LMS also provide synchronous (e.g., chats) and asynchronous (e.g., message boards) communication tools to make up for the lack of physical contact between students and teachers, as well as among students, in order to make social construction of knowledge possible. As learning becomes ubiquitous—learning and interactions may happen anywhere, anytime—message boards (forums) become an essential part of social learning in online environments.

In collaborative, project-based and teamwork-based learning, social learning is at the center of the process. Social learning builds on the notions that cognitive processes take place in a social context, by reciprocal interaction between behavior and controlling conditions, both individual and environmental (Bandura, 1971), and that knowledge is created and constructed by the interactions of individuals within society (Berger & Luckman, 1967). In addition, research on CSCL and virtual communities of practice has also shown interest in knowledge creation by participation and engagement in the discourse (Hmelo-Silver & Barrows, 2008; Lave & Wenger, 1991; Zhao & Chan, 2014).

In formal online learning contexts, the interactions, participation, social exchange, and discourse-based knowledge building processes happen essentially in course forums. Therefore, it is only natural that an important stream of research has focused on describing, explaining, and understanding the social dynamics that take place in forums on online courses. One of the most novel approaches to the study of social dynamics in online courses is the application of SNA to course data, known as social learning analytics (e.g., Oshima, Oshima, & Matsuzawa, 2012).

According to Buckingham-Shum and Ferguson (2012), inherently social learning analytics has two different aspects: social network analytics and discourse analytics. The former focuses on SNA of course data in order to explain and understand the social dynamics of the course, and it will be the main focus of this chapter (i.e., we shall restrict the concept of social learning analytics to SNA of educational data); the latter explores the nature of the contents and structure of the discourse between learning agents in a course, which is out of the scope of this study. Buckingham-Shum and Ferguson state that the underlying idea behind social learning analytics is that networked learning supported by ICT consists of actors (both people and resources) and the relations between them, and that social network analysis investigates these network processes and the properties of ties, relations, roles, and network formations. Therefore, social network analysis brings together graph theory and sociology and communication to improve learning processes. The main uses of social learning analytics include detection of communities (Buckingham-Shum & Ferguson, 2012) and identification of relevant learning agents, such as at-risk students, knowledge brokers, or influential students (Hernández-García, González-González, Jiménez-Zarco, & Chaparro-Peláez, 2015).

Social learning analytics facilitates this identification in two ways: analysis and visualization. Analysis focuses on calculation of SNA parameters and metrics for each node—see Freeman (1979) for further information about centrality measures and Hernández-García (2014; p. 156) for SNA metrics and indicators for learning analytics—and network overall parameters, such as average network degree (average number of incoming, outgoing, or global links of a node in the network), network density (the number of total edges present in the network relative to the number of edges in a full-connected network), or network diameter (the largest number of nodes that must be traversed in order to travel from one node to another).

Visualization of social networks facilitates the identification, at a glance, of students who are disconnected from the network; furthermore, filtering and visual transformations of the graph, based on relevant metrics or node attributes, may help understanding the social dynamics of the course. The main advantage of the analysis is that it also provides a numerical way to characterize different aspects of the social graph (although the meaning of the different SNA parameters may be difficult to understand for instructors with no knowledge of SNA, and therefore the usefulness of the analysis is limited to the subject's ability to interpret the results). Social graph visualizations complement the analysis in a direct and eye-candy way, once the main concepts are learnt.

Regarding social learning networks, Hernández-García et al. (2015) state that teachers usually have access to one part—the visible one—of the social exchanges and participation in a course. Messages posted to the course message boards represent this visible part. More often than not, assessment in online courses rely on the final evidence from quizzes or essays that students deliver in the learning platforms, but also on evaluation of students' participation and quality of content posted to forums. Furthermore, this visible activity also allows instructors to determine whether the different concepts are actually being learnt by students and to detect lack of active engagement in the discourse. Nonetheless, Hernández-García et al. claim that there is another type of passive social exchanges where individuals interact not with teachers and other students, but with the content created by others, and that this kind of interaction that may pass unnoticed to instructors can provide additional information about student engagement. According to Wise and Hausknecht (2013), the lack of active engagement in conversations does not mean a lack of involvement or that learning is not happening, because students have different learning styles, and some students may enhance their learning with external knowledge that they do not share, or may act as learning witnesses or "invisible students" (Beaudoin, 2002), and build their learning around content created or shared by others.

Hernández-García (2014) proposes the suitability of SNA tools to perform social network analytics of both types of networks. In order to do so, he shows some examples of use of Gephi for SNA, by using data from a proprietary learning platform. Hernández-García divides forum log data into three different datasets: relations among users based on their posting behavior (the "reply network"; i.e., who replies to whom), relations among users based on message viewing behaviors (the "read network"; i.e., who reads messages posted by whom), and relations among messages (a network that relates each message to its parent in a discussion, displaying threads as message trees).

The objective of GraphFES, the Web service presented in this chapter, is to automatically build these three networks from LMS data logs (more specifically, data from Moodle logs) and show an example of the potential of social learning analytics with data from two courses with different characteristics.

Social Learning Analytics: Tools

This section offers an overview of existing tools available for social network analytics in formal learning environments (VLEs and LMS). The analysis will focus on SNA tools for Moodle, the leading open-source LMS (Edutechnica, 2015) and will detail three tools oriented toward social learning analytics (Social Networks

Adapting Pedagogical Practice (SNAPP), Forum Graph and Meerkat-ED), as well as generic SNA software for social learning analytics. Section "Case Studies" will show the different visualizations provided by each tool.

Social Networks Adapting Pedagogical Practice

SNAPP[1] is a web browser bookmarklet that extracts information about message board activity from the most widely adopted LMS (Sakai, Blackboard, Moodle, and Desire2Learn), and then builds up the resulting social network in a Java applet. The two existing versions of SNAPP (1.5 and 2.1) have similar functionalities.

SNAPP's Java applet shows different tabs, the first three of which are interactive. The first tab shows the social network graph and allows manipulating it by filtering, applying different layouts and selecting individual nodes—nodes in SNAPP represent participants in the message board. SNAPP 2.1 also displays a timeline of the messages posted in the forum. The second tab shows each user's number of posts in SNAPP 1.5 and the main social network parameters (degree, in- and out-degree, betweenness and eigenvector centrality, and network density) in SNAPP 2.1. The third tab allows exporting the graph in GraphML and VNA formats in SNAPP 1.5, or writing annotations in SNAPP 2.1 (export capabilities are included in the first tab in SNAPP 2.1., in addition to the ability to export to Gephi's GEFX format).

Lack of applet updating causes SNAPP 2.1 to not work properly in latest versions of Moodle. Neither versions of SNAPP could be tested with the courses data for comparison. Furthermore, proper installation requires configuration of security exceptions in the Java Runtime Environment and connection to an external source to perform the analysis. The process of social graph construction includes loading and rendering of all the threads and posts in a message board, and parsing and processing of the HTML content.

More information on publications covering the use of SNAPP can be found at http://www.snappvis.org/?page_id=20

Forum Graph

Forum Graph[2] is provided as a report plug-in in Moodle's repository and creates the social graph of one single forum. The visualization of the resulting social graph only displays one possible representation of data, with node sizes representing user's number of posts, and edges representing the number of times a user replies to another user. The social graph can be exported as an SVG image. Additional information includes a tooltip showing the number of discussions initiated by each

[1] http://www.snappvis.org

[2] https://moodle.org/plugins/view/report_forumgraph

user, the number of replies a user has made, different colors for teachers and students, and direct access to each user's log in Moodle's Legacy Log (see section "Moodle Logs and Data Extraction and Visualization") by clicking on them. The plug-in also shows a list with the three top contributors to the forum. Despite its ease of installation, the visualization options of Forum Graph are very limited and may not be suitable for courses with high number of students (due to display size limitations). Furthermore, Forum Graph does not include any SNA tools or information about the main SNA parameters.

Meerkat-ED

Meerkat-ED[3] is a Java application developed by Reihaneh Rabbany that loads information about forums and posts from a Moodle backup file (.xml and .mbz files, depending on the version of Moodle; its use therefore requires that the user has course backup/restore permissions and a backup of the course), extracts the information, and then constructs the social graph. Meerkat-ED includes both social network analytics and discourse analytics capabilities.

Regarding social network analytics, Meerkat-ED gives information about students' posting activity (i.e., the "reply network" in Hernández-García et al., 2015) and their degree centrality, as well as basic modularity information (weakly connected components which indicate the existence of different communities). It also shows an additional graph that represents centrality over a target, with more central users nearest to the center of the target. Meerkat-ED provides basic node manipulation (dragging and selecting nodes, dragging the network and zooming). An interesting characteristic of Meerkat-ED is that it allows dynamic analysis of interactions by selecting the timespan and dragging a timeline.

As for discourse analytics capabilities, Meerkat-ED allows filtering by forum, discussion, and posts, and builds a network with the most used terms and their relations. Graphically, it shows all thread titles in nested mode, a graph depicting the relations between terms, a table with the number of occurrences of each term, and a cloud of the different terms.

SNA Tools

The main problem with built-in plug-ins like SNAPP and Forum Graph is that they provide little information other than visualization of the network topology, and therefore they are very limited in terms of social network analytics capabilities. Meerkat-ED can be considered an intermediate step that shows how external apps can improve analysis and visualization by separating the data layer from Moodle

[3] http://webdocs.cs.ualberta.ca/~rabbanyk/MeerkatED

logs and the process and presentation layer done in the application, with the added value of basic discourse analytics capabilities.

Nevertheless, although they may be suitable to view basic information about the social interactions that take place in forums on an LMS, these tools lack the advanced SNA capabilities and advanced graph interaction and filtering that are necessary for in-depth analysis and understanding about the social learning happening in a course. Furthermore, the three tools presented in this section allow users to observe the visible networks but, despite that data being available in the LMS, none of them provides any information about the invisible network of forum reading activity.

SNA software tools, on the other hand, are specifically designed to perform these tasks. There are many proprietary and open-source solutions available for general SNA that can help carrying out social learning analytics. However, despite their suitability for SNA, these systems also have some disadvantages:

- Because they are general purpose SNA applications, they may require some adaptation for social learning analytics purposes.
- Their functionality is restricted to the domain of SNA, and therefore their use may require some training for effective analysis. Moreover, while the concepts involved in the analysis are the same, the operation of each tool may be completely different from one application to another.
- Data from LMS and formal learning systems is stored in formats that are exclusive to each platform, and generally the design of the databases is not ready for SNA. Therefore, SNA of educational data from online platforms usually requires data extraction from LMS databases, and processing and transformation of the extracted data to a format readable by SNA programs.

Some authors advocate for the use of SNA tools for social learning analytics, but they also pinpoint the need for development of plug-ins that may translate the data from LMS to SNA applications (Amo Filvà, García-Peñalvo, & Alier Forment, 2014; Hernández-García, 2014). This study aims to cover this gap in the case of the open-source LMS platform Moodle by introducing GraphFES.

The main objective of GraphFES is to provide a data extraction layer that transforms data from Moodle to Gephi—an SNA program—for social learning analytics. Understanding the process of data transformation in GraphFES requires to study the data source system (Moodle logs), the data target system (Gephi) and the design and operation of the transformation tool (GraphFES).

Moodle Logs and Data Extraction and Visualization

Moodle has a built-in logging system that stores every user interaction with the LMS. LMS logging systems are a critical source of information for the purposes of analysis, study, and visualization of interactions—and, more specifically, the social interactions that take place in online education.

Despite registering all the learning platform's activity, earlier versions of Moodle did not retrieve enough information about the learning contexts of these interactions in the system logs, and they had performance and scalability issues. Driven by the emergence of learning analytics, Moodle version 2.6 introduced an enhanced version of the logging system that facilitates different kinds of analytics.

Moodle's new logging system has many benefits when compared to the legacy log system (Moodle, 2015). It captures richer information, abstracts log reading and writing for higher scalability, monitors gathered information and facilitates storage in external systems for analysis and visualization. Therefore, from version 2.6 onward there are three logging systems in Moodle: the new version, known as Standard Log; the old version, known as Legacy Log; and the External Log, which allows connection to an external log database.

Moodle uses the Events API and the Logging API to generate and store logging information. The Events API provides a notification and a unique event collection system for the different actions that users can perform in the LMS. The Logging API consists of different plug-ins for configuration, registration, and reporting of data triggered by the different events.

When a user performs an action in any module in Moodle, the system generates an event. The log manager listens to events and, depending on system configuration, determines whether to register and log the event or not. If the event must be registered, then the log manager passes the information to the plug-ins, and they store the information in the corresponding database table.

Data extraction, reporting and visualization, on the other hand, requires reading data from the tables storing that information. For example, Moodle's built-in activity report makes a query to the log manager to verify what kind of logs it can read, including data source selection when there is more than one source available. When the activity report module is granted access, it looks up the registered events and shows them on screen.

GraphFES is a web service and application that allows external queries to Moodle's Standard and Legacy logs. GraphFES is thereby a tool for generation of social graphs with data from forum activity in Moodle. Section "GraphFES: Design and Operation" later details how the design of such a tool requires development of a Moodle local extension implementing functions for external data extraction from requests to Moodle's log systems via a web service that uses the REST protocol. GraphFES, unlike SNAPP or Forum Graph, does not allow direct visualization of data in Moodle, and uses Gephi for data analysis and visualization. The next section presents Gephi and GEXF, Gephi's data format to represent and analyze social graphs.

Gephi: A Tool for Social Network Analysis

Data Format and Dataset Characteristics

As mentioned earlier, Gephi will handle the analysis and visualization of the social networks from interactions in Moodle message boards. The reason for the choice of Gephi is that it is a widely used open-source software program, with continued

Fig. 11.1 Comparison of formats supported in Gephi and their features (Gephi, 2015)

support and an active community. Furthermore, Gephi is oriented toward generic graph and social network analysis, and it can be easily extended to suit users' needs by installing NetBeans plug-ins. Gephi currently supports the following data formats (Gephi, 2015):

- GEXF (Graph Exchange XML Format)
- GUESS's GDF
- GML (Graph Modeling Language)
- GraphML (Graph Markup Language)
- Pajek's NET
- GraphViz DOT
- CSV (Comma Separated Variables)
- UCINET's DL
- Tulip's TPL
- Netdraw's VNA
- Spreadsheet (MS Excel and other programs)

The choice of the most adequate data format for GraphFES's output requires an analysis of Gephi's functionalities. Figure 11.1 shows a comparative table of the features of the different graph formats supported by Gephi. From Fig. 11.1, it is evident that GEXF has more features than the rest of formats. Furthermore, while other alternative and popular formats (UCINET DL, Pajek NET, GML, Netdraw VNA) are also compatible and supported by Gephi, their cross-compatibility is not as good (e.g., Pajek does not allow the use of attributes, and therefore it is only useful for analysis and visualization of network topologies). On the other hand, a complete analysis of the social graphs from course forums requires being able to collect additional data and incorporate them as extra information about nodes and edges. GEXF allows storing of this additional data as nodes' and edges' attributes, and therefore the choice of GEXF as data format is most likely the best fit to the characteristics of GraphFES.

Because it is XML (eXtensible Markup Language), GEXF is a consolidated, extensible, and open format. Another advantage derived from being an XML is that there are XML parsers available for all programming languages, allowing developers to process a GEXF file on practically any kind of application, regardless of the programming language or operating system it is coded in.

A GEXF definition of a graph consists of nodes, edges, and the data associated with them (GEXF Working Group, 2015). A very simple example of a graph in GEFX is the following:

Ex. 11.1 GEXF document of a simple graph

```
<?xml version="1.0" encoding="UTF-8"?>
    <gexf xmlns="http://www.gexf.net/1.2draft" version="1.2">
        <graph defaultedgetype="directed">
            <attributes class="node">
                <attribute id="0" title="username" type="string"/>
            </attributes>
            <nodes>
                <node id="0" label="NodeA">
                    <attvalues>
                        <attvalue for="0" value="John"/>
                    </attvalues>
                </node>
                <node id="1" label="NodeB">
                    <attvalues>
                        <attvalue for="0" value="Mark"/>
                    </attvalues>
                </node>
            </nodes>
            <edges>
                <edge id="0" source="0" target="1"/>
            </edges>
        </graph>
    </gexf>
```

The XML document above consists of a declaration of the document as GEXF (identified by its namespace), followed by the additional attributes for nodes and edges (in this case, only one additional attribute for nodes, with attribute id equal to 0 and attribute name equal to “username”). Then, the document lists the network

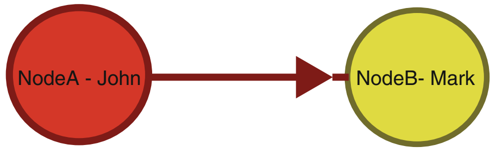

Fig. 11.2 Visual representation in Gephi of the graph from the GEXF file in Example 11.1

nodes (they include a node id and a label), including their attributes (in the example, the node with id = 0 would have username="John" as attribute, and the username of the node with id = 1 would be "Mark"). After all nodes have been listed, they are connected by declaring the edges, with their respective source and target nodes (the example only shows one edge connecting nodes 0 (labeled "NodeA" and with username "John") and 1 ("NodeB," "Mark"). Graphically, the visual representation of this document in Gephi would be the graph shown in Fig. 11.2.

GraphFES: Design and Operation

GraphFES as Web Service

GraphFES (Graph Forum Extraction Service) comprises two different elements: a local Moodle extension and a web application, and they serve two different purposes: data extraction and social graph building, respectively. Raw data extraction from the Moodle log tables requires the implementation of the functions that will be accessible via the web service and therefore needs to be managed by the local extension. The web application, on the other hand, serves as front-end and makes the requests to the web service in order to generate the different types of social graphs with the data it receives.

The local extension is programmed in PHP language (the same as Moodle), and its design follows the template for web service creation in Moodle. The local extension is therefore installed as a plug-in that implements two external functions (*forum_reportAllLegacy* and *forum_reportAll*). The reason to implement two functions instead of one is to ensure compatibility with both Moodle's legacy log table (*mdl_log*) and the new log table (*mdl_logstore_standard_log*). This guarantees that the web service may also be able to extract data from imported Moodle courses from versions 2.6 and lower that include log data. More specifically, the database queries made by these two functions are the following:

```
SELECT * FROM mdl_log WHERE module="forum" AND course=$courseids[0]
SELECT * FROM mdl_logstore_standard_log WHERE component="mod_
forum" AND courseid=$courseids[0]
```

The implementation of these two external functions in the internal plug-in makes the log data accessible to a web service owing to Moodle's Web services Application Program Interface (API). In order to access the data required to build the social graphs from Moodle logs, two additional operations must be performed in the

Table 11.1 Functions needed to create the web service

Function	Description
core_enrol_get_enrolled_users	Gets enrolled users by course id
core_course_get_courses	Returns course details
mod_forum_get_forums_by_courses	Returns a list of forum instances in a provided set of courses
mod_forum_get_forum_discussion_posts	Returns a list of forum posts for a discussion
local_graphfes_forum_reportAll	Full forum report
local_graphfes_forum_reportAllLegacy	Full forum report from legacy logs

Moodle platform: habilitation of the web service and activation of the REST protocol. These options appear in Moodle's administration menu, under the section "External services." In order to create the web service, it is necessary to create a new service with the functions indicated in Table 11.1.

The above operation activates the web service, but a complete setup also requires to manage access authorization. Administrators can grant authorization to individual users for using the web service in the "External services" menu, and they can then generate access tokens—if required—for authorized users in the "Manage tokens" section. In this version of GraphFES, generation of tokens is not necessary because the web application manages the authentication process after input of user login data (username and password).

The web application that serves as front-end and that builds the social graph is programmed in Node.js and the Express web framework. The combination of this programming language and the framework speeds up the development process and makes it simpler due to the high-speed of Node.js, which uses Google's JavaScript 8 engine. The use of JavaScript also allows developers to include different open-source libraries that facilitate the creation of the social graphs.

The main reason for the choice of a web application as front-end is that it facilitates the implementation of the application in any server, making it possible to access the application remotely with any browser. Additionally, it can also run locally in any computer. Besides, since the web application is programmed in Node.js, it is compatible with the most popular operating systems (Windows, Mac OS X and Linux, among others).

The structure followed by the web application follows the default generation structure of the Express framework. Apart from that, and because the role of the web application is not to store data but to request, structure, and transform them to a GEXF file in the faster and more effective way, no additional database is required.

Using GraphFES

This section details the design and operation of the web application and how it generates the different social graphs, step by step, and in a simple way. Once the application is loaded and the server is waiting for requests (users can do this locally by

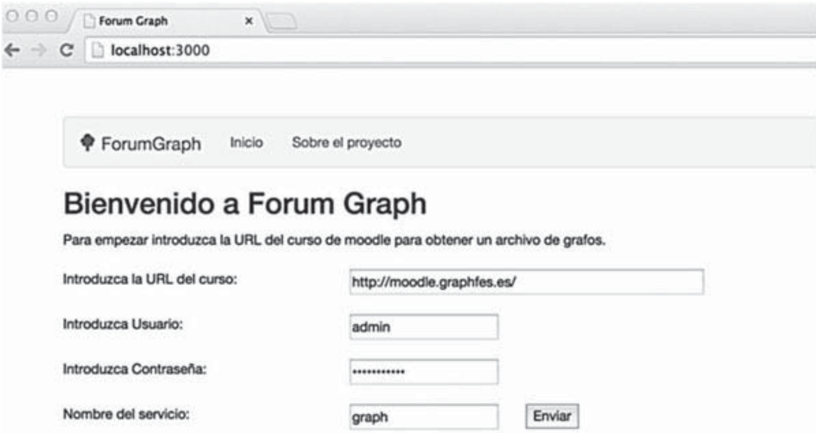

Fig. 11.3 Application main screen

running the main JavaScript file, app.js, with node from the command line), the application is accessible using any web browser (if run locally, the default access URL is http://localhost:3000). The main front-end screen asks the user to complete a form (Fig. 11.3). The different form fields are: URL of the Moodle platform, username, and password, and the name given to the web service in Moodle.

Upon introduction of the values for each field and form submission, the application sends a login request to the Moodle LMS. If successful, Moodle sends a token (the application will use this token for the different REST calls for data extraction) and the browser will redirect the user to a new screen for course selection that lists the courses to which the user has access. The data extraction process and the graphs generation start after selection of a course.

The process begins with a request to the function *core_enrol_get_enrolled_users* in order to retrieve the list of students enrolled in the course (this is important in order to also receive data from students who have no activity in the course) and another request to the function *mod_forum_get_forums_by_courses*. This function retrieves all the existing message boards in the course.

After having received this information, the application does the request for Moodle log forum data extraction. This request is done using the two functions implemented by the local extension (*local_graphfes_forum_reportAll* and *local_graphfes_forum_reportAllLegacy*) and it only retrieves log data related to forum activity. The application then differentiates between activities associated to discussion/post-creation and discussion/post-views. Additional data about users, posts, and discussions (e.g., user id, message content, timestamp) are also temporarily stored in memory in order to include richer information in the social graphs.

With all the different data, the application uses an open-source library (*element-tree*) that creates an XML document with three different social graphs, in a similar fashion to the datasets in Hernández-García (2014). The XML documents are already formatted in GEXF format and are stored in the graph folder of the application. The three different graphs correspond to the following files:

- *Views.gexf*: this graph shows the relations of messages viewed by course participants. In other words, it provides information about how many times user *a* has read a message posted by user *b*.
- *Replies.gexf*: it shows the connection between students based on who replies to whom, and how frequently.
- *Messages.gexf*: the graph shows the connection among messages (i.e., which message is a reply to another message).

In the Views and Replies graphs, each node corresponds to a course participant. Nodes in the Views graph includes information about user id and username as node attributes, while nodes in the Replies graph have additional information about user id, username, number of total posts, number of initial posts in a thread, and number of replies.

On the other hand, the Messages graph considers that each node is a message posted in one of the courses' forums. As with the two other two graphs, each node has additional information—as attributes—about:

- Name and id of the forum it was posted to.
- Post title and id.
- Message content.
- Post timestamp.
- Author's name and id.

Case Studies

In order to test the operation of GraphFES, data from two different courses were extracted from Moodle version 2.8.3. The original course data were collected from two different Moodle installations with versions lower than 2.6 and were then anonymized and restored to the Moodle 2.8.3 used to test the web service. Therefore, the original data were stored in the Legacy log, and activation of the Legacy log in the Moodle 2.8.3 was necessary. The following sections give an overview of the two courses and the results of the empirical analysis.

Context and Description of the Courses

The two courses chosen for the study were one online course from the Online Master's in Domotics and Digital Home, and an undergraduate programming course for first-year students of the Biotechnology degree at Universidad Politécnica de Madrid. There are two reasons for this choice of courses: first, there are many differences among them in terms of number of students, duration, forum use intensity, methodology, and instructional goals; the second reason is that the relatively low forum activity of the first course allowed us to easily and quickly test and compare that the output from GraphFES was correct, while the high forum activity in the second course allowed us to test the scalability of the web service (in addition, some course data was selected to check correct operation of GraphFES in this course, too). This data analysis will cover both courses, but it will focus primarily on the programming course due to the higher complexity of the resulting networks.

The online Master's course ("Socioeconomic analysis of the domotics and digital home environment," Course 1) is the first of eight mandatory courses of the Master's degree. Most of the students are architects and electrical and telecommunication engineers (i.e., there is an overall mixed background with regard to the use of information technologies). The course comprises two different modules, with 14 enrolled Spanish and South American students—in different locations—and a duration of 2 weeks. During those 2 weeks, students have access to the lectures' contents and to supplementary information and links, and they have to complete two quizzes and an individual and a group assignment based on a case study. Three different groups were formed for the group case study. There are five forums available for interaction (one debate and one teaching support forum per module, and a forum for the group assignment where students may only access and use their group's threads). Apart from the group assignment, students are expected to work individually.

The undergraduate course is "Programming Basics" (Course 2), a one-semester long mandatory course in the Biotechnology degree, with 110 students living in the Madrid area and that presumably share other in-class courses. Although there is an in-class two-hour-long introductory session, the whole course is based on project-based online teamwork, following the Comprehensive Training Model of the Teamwork Competence (CTMTC) (Leris, Fidalgo, & Sein-Echaluce, 2014).

The CTMTC has a strong focus on teamwork, and therefore students were distributed in 19 groups (with an average of six members in each team, a minimum of five members and a maximum of seven). The CTMTC determines five phases of the project, with three types of evidence for assessment along them: individual, group, and results (Fidalgo-Blanco et al., 2015). Assessment of the group and results is based on contributions to wikis and file-sharing services (e.g., Dropbox), upon which a grade is given to each group. The use of forums (there are Q&A, teaching support, and group forums) is a critical part of the instructional method because assessment of individual evidences is based on participation and contributions to the group forums. Students in this course are expected to build knowledge by working together on a project as a team.

Data Extraction

The total number of records related to forum activity is 1850 and 80185 for the first and second course, respectively. Processing times of GraphFES are almost instant for Course 1, and under 1 min for Course 2 (for comparison, Meerkat-ED cannot complete graph computation of Course 2). As expected, creation of the Views graph (the invisible or read network) takes most of this processing time because the

Fig. 11.4 Course 1: Initial resulting networks of Views (*left*), Replies (*center*), and Messages (*right*)

Fig. 11.5 Course 2: Initial resulting networks of Views (*left*), Replies (*center*), and Messages (*right*)

number of edges (relations) may grow exponentially as the number of posts increases—the number of edges in the Replies and Messages graphs is proportional to the number of users and messages posted, respectively.

After both sets of graphs are generated by GraphFES, data is ready for analysis in Gephi.

Initial Data Analysis

Figures 11.4 and 11.5 show the initial visualization of the resulting networks of Views, Replies, and Messages in Gephi after application of a Force Atlas 2 data transformation to Courses 1 and 2. Note that Fig. 11.4 shows a Radial Axis transformation of the Messages graph (right), as suggested by Hernández-García (2014), but in Fig. 11.5 (right) we use Force Atlas 2 due to the high number of nodes—Gephi cannot display more than 128 root nodes in the Radial Axis visualization. For comparison, Fig. 11.6 shows the Replies network of the two most active forums in courses 1 and 2 from Forum Graph, and Fig. 11.7 shows the Replies and Messages network of course 1 from Meerkat-ED.

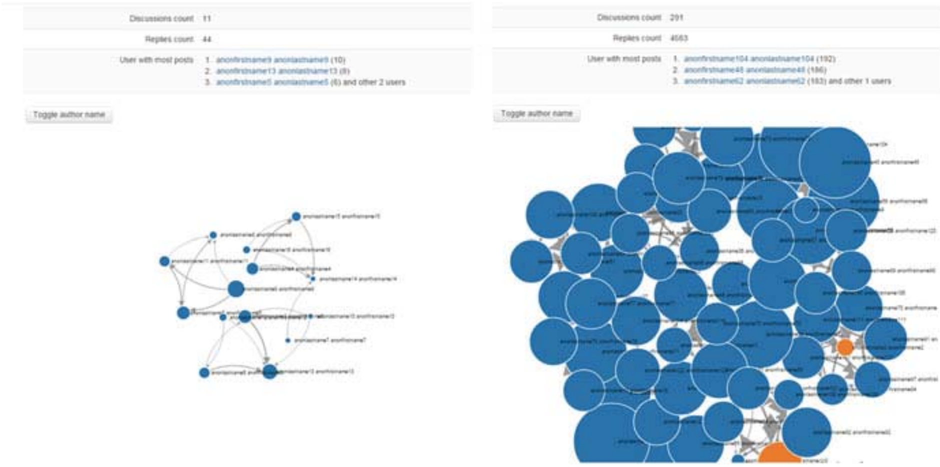

Fig. 11.6 Visualization of the Replies graph of the most active forum in courses 1 (*left*) and 2 (*right*) in Forum Graph

The visualizations depicted in Figs. 11.4 and 11.5 give some evident information about the different courses:

- As expected, the Views networks have a much higher number of connections than the Replies networks (passive versus active participation).
- The connections among students in course 2 reflect primarily intra-group communication (indicating that group members focus on the teamwork project), while course 1 shows more diverse exchanges among students (balancing individual and group-based learning).
- The Views and Replies networks in course 2 allow detection of isolated/disconnected students, who are not participating actively (Replies network) and/or passively (Views network) in the course.
- The Messages graph of course 1 shows which are the most active threads and posts within a thread, but there are simply too many messages in course 2 to perform a visual analysis.

Additionally, in the top-right part of the main window (not shown in Figs. 11.4 and 11.5), Gephi gives information about the number of nodes (users in Views and Replies networks and posts in the Messages graph) and edges (existing relations between nodes).

From Figs. 11.4, 11.5, 11.6, and 11.7, apparently Forum Graph and Meerkat-ED offer additional information when compared to Gephi, at least in course 1. However, none of the former two really provide much useful information about forum activity in course 2 (Meerkat-ED fails to load the course, and the graph nodes in Forum Graph are too cramped to extract any useful information). Furthermore, centrality values in Meerkat-ED are not weighed, and therefore it is difficult to retrieve information about which students are participating the most.

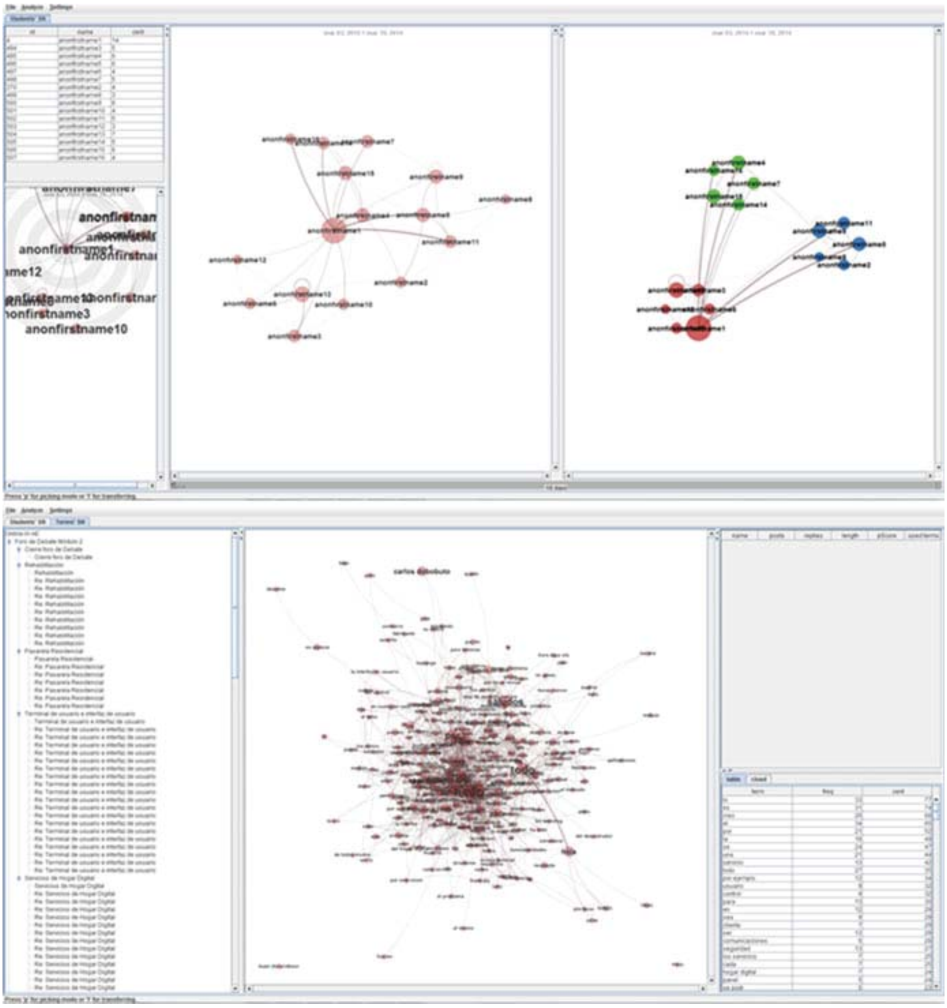

Fig. 11.7 Visualization of the Replies (*top*) and words in messages (*down*) of course 1 in Meerkat-ED

From the above, it would seem that existing tools have important limitations in order to perform successful and effective social learning analytics. However, this is where the additional features of specific SNA tools like Gephi shine and cover this gap, expanding social learning analytics capabilities. The most important features available in Gephi for these purposes, which allow creating more informative visualizations and will be covered in the following section, are:

- Calculation of SNA parameters
- Nodes' and edges' partitioning and ranking
- Filtering

Social Learning Analytics in Gephi

Calculation of SNA metrics and parameters—as in section "Social Learning Analytics," we refer to Hernández-García (2014; p. 156) for further information—is an essential part of social learning analytics for three reasons. First, the values from the analysis provide meaningful information that can be directly interpreted in terms of student participation, passive and active engagement (e.g., node centrality, edge weights in the Views and Replies networks), knowledge brokerage (e.g., betweenness centrality), leadership, authority or expertise (e.g., authority, pagerank, eigenvalue centrality), information collectors or hubs, and overall network information such as cohesion (e.g., density) or identification of communities (e.g., connected components, modularity, clustering). Second, the results of each analysis are incorporated to Gephi's data laboratory; this means that all the SNA parameters calculated are incorporated to each node or edge in a data table, which can later be exported for further non-SNA statistical analysis (such as multiple regression analysis or structural equation modeling) in other statistical software applications. Individual node parameters are also available in the information window (in a tab named "Edit") when a node is selected in the graph. Third, output variables of the SNA are added as variables to the main panel, and they become available for partitioning, ranking, and filtering.

Partitioning and ranking facilitate adaptation of the different network visualizations to the users' needs, by emphasizing aspects that the observer may consider of most interest. Partitioning assigns different colors to nodes or edges that share the same values of a given SNA parameter or node/edge attribute. Since GraphFES includes additional information about students and messages as node attributes, that information becomes already available for partition purposes, too (e.g., we could assign different colors to students with the same number of initial posts, replies or total posts, or to messages posted in the same forum or by the same user). Moreover, partitioning gives information about the percentage of nodes that are included in each partition.

Ranking is one of the most interesting features of Gephi, and it allows assigning different sizes and colors to nodes and edges, in adjustable scales, depending on the values of the chosen SNA parameter or attribute, or just to a range of them. Ranking is extremely useful because it gives a direct visual interpretation of the aspects of interest, both in absolute and relative terms. For example, ranking node size by weighed out-degree and node color by weighed in-degree on the Replies graph would provide information about who has written more (or less) posts and who has been replied most or least. Additionally, users can select whether they want to label nodes and edges with any SNA parameter and/or attribute.

Figure 11.8 shows how the selection of SNA parameters and attributes affects overall network configuration. On the left, node size ranking uses weighed out-degree, node color ranking uses weighed in-degree and nodes are labeled with the username. On the right, the only change is node size ranking criterion (betweenness centrality). The figure shows that, despite having written very few posts (small

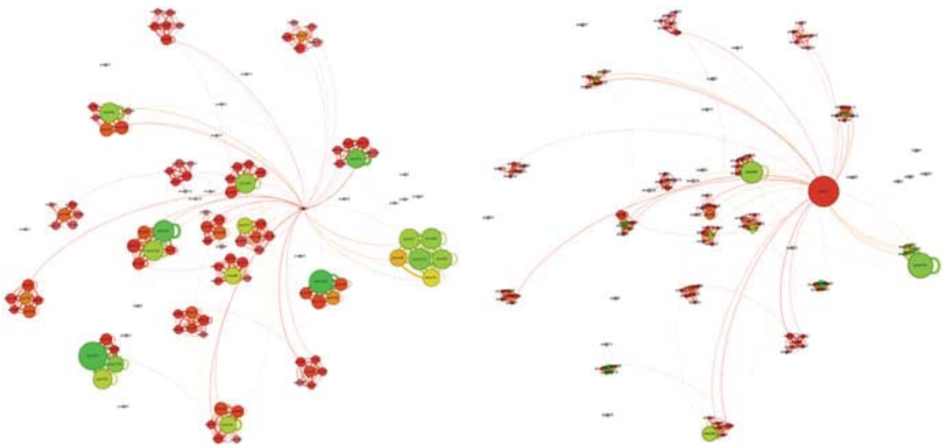

Fig. 11.8 Replies graph (course 2). Node size ranking by weighed out-degree (*left*) and betweenness centrality (*right*). Node color ranking by weighed in-degree (*grey* equals 0, higher in-degree from *red* to *green*)

size on the left) and received relatively few replies (red color on both), the teacher (user anon3) plays the main bridge or information broker in the course.

Filtering is also a powerful feature of Gephi. Filters are available as a tab in the main right panel. As said earlier, after computing the values of SNA metrics and parameters, Gephi makes them available for filtering—attributes are also initially available for filtering purposes.

The inclusion of attributes and SNA parameters for filtering vastly enhances the usefulness of SNA software programs like Gephi. By specifying different filters, users can visualize only relevant parts of the network, or analyze again the resulting networks including only the nodes or edges that fulfill the specified conditions. Depending on the type of variable, users have the option to apply many types of filters (partition, range, logical operations, dynamic processes, topology-related aspects such as levels of ego-networks, and even semantic web analysis via SPARQL queries if the attributes include semantic information). Interestingly, when a filter is applied, the data laboratory only shows information about the nodes and edges affected by the filter. Additionally, users can save simple and complex filters for later reuse.

While setting the appropriate filters for each class and instructional method may not be straightforward, some types of simple filters may provide lots of useful and actionable information. For example, in networks with high number of nodes, a partitioning filter of the Views and Replies graphs shows the students that have not read or written any messages (out-degree equal to zero, Fig. 11.9, left) or, in the Messages graph, which messages have not been answered yet (out- and in-degree equal to zero).

Filtering can be applied in successive stages. For example, after identification and selection of a potentially low-connected student, anon83 (Fig. 11.9, right), immediate node information is available in the upper left side. The student has in-

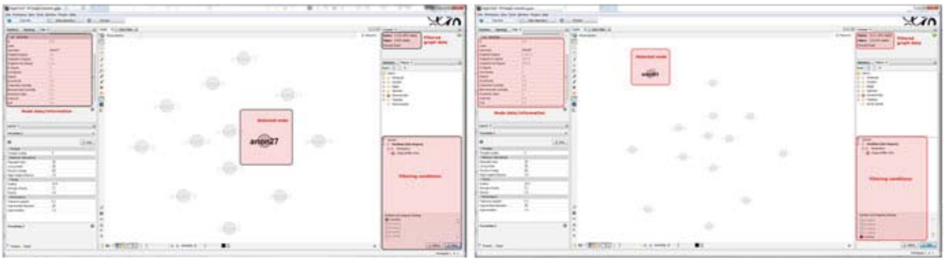

Fig. 11.9 Disconnected (*left*) and disconnected and low-connected (*right*) students in the Views graph

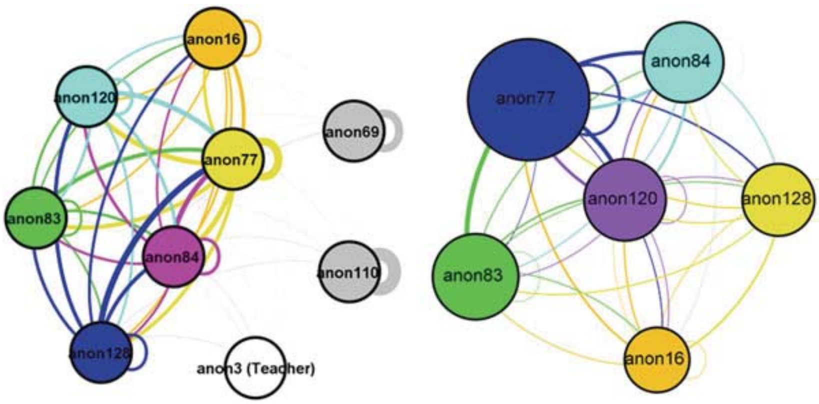

Fig. 11.10 Views (*left*) and Replies (*right*) first level ego networks of student *anon83*

degree of 7 and out-degree of 9. That means that he has only read messages from other nine people; he has read 5413 messages—above the network average—and his messages have been read 6077 times. Note that due to a limitation of Moodle logs, whenever a discussion is viewed, GraphFES considers that all the messages in the thread are viewed. Moreover, there is also information about node id; based on this information, the user can perform multiple other actions, such as further filtering to show the student's ego network, both in the Views and Replies graphs, to show whom the student is connected to (Fig. 11.10).

From Fig. 11.10, we observe that student anon83 has actually only read messages from other seven students (the other five students in his team and two students that do not belong to his team), the teacher and his own posts. Furthermore, edge thickness show that his reading has been focused on messages written by his teammates.

The above is just a simple example of how the use of SNA tools for visualization of learning networks can provide further insight and information about the dynamics of online learning. The example shows how social learning analytics and visualization of learning networks help identifying disconnected students, but also how it may offer additional information about learning agents (Figs. 11.9 and 11.10 show that student anon83, a potentially disconnected student, is just not engaged at course level but he is actively participating in the teamwork).

There are many other combinations of ranking and filtering from SNA that offer relevant information about the different social aspects of learning in online courses. For instance, we could have focused on identification of central learning agents (e.g., by making a deeper analysis of Fig. 11.8) or building of learning communities within the course (by analyzing modularity and weakly connected components of the networks, for example).

Nevertheless, and due to length limitation, the objective of this chapter is not an in-depth exploration of each of the possible uses of Gephi's features for social network analytics, but rather to show the potential of SNA tools to perform this type of analysis. Of course, the type of filters and analyses in Gephi with data from GraphFES should be tailored to the institution's and teachers' needs, taking into account that the type of course and instructional method also affect how the results may be interpreted.

This means that an institution-wide plan regarding social learning analytics strategies should be deployed in order to effectively define what ranking, partitioning, and filtering might be most useful for analysis, or what SNA parameter values might be used as learning indicators for successful learning.

Although we have already mentioned how Gephi may help extracting SNA parameters for ranking, partitioning, and filtering purposes, so far we have focused on the visual aspects of SNA tools for social learning analytics, paying little attention to the information that is directly available from SNA. The next section will give a brief outline of the SNA results for courses 1 and 2, and some possible interpretation from the values of SNA parameters.

SNA Parameters

Gephi incorporates different types of calculations of SNA metrics. Most analyses in Gephi occur in three steps: running the analysis (some additional parameters may be needed, such as specifying directed or undirected networks), HTML report of results (the reports generally include the main results and some graphics of distributions of SNA parameter values), and incorporation of the values to the data laboratory (and availability of SNA metrics for partitioning, ranking, and filtering). Both overall network and individual node SNA parameters can be calculated in the same operation (e.g., calculation of network diameter or average path length entails calculation of each node's eccentricity, and betweenness and closeness centralities). Table 11.2 gives an overview of the most relevant network parameters of Views, Replies, and Messages graphs of courses 1 and 2.

The parameters of the Views graph have useful information for instructors about the "invisible" network. In online learning, the average degree of the Views graph should ideally be as close as possible to the number of nodes in the network (this is the same as a network density of 1 or an average path length of 1). That would mean

Table 11.2 Main overall network metrics

	SNA metric	Views	Replies	Messages
Course 1	Nodes/edges	16/222	16/64	153/123
	Av. degree	13.88	4	0.8
	Av. weighed degree	348.31	7.69	0.8
	Av. path length	1.08	1.94	1.58
	Diameter	2	4	5
	Density	0.93	0.27	0.01
	Modularity	0.15	0.32	0.89
	Number of communities	2	2	32
	Weakly connected components	1	1	30
	Strongly connected components	2	1	153
Course 2	Nodes/edges	123/2854	123/662	9241/8604
	Av. degree	23.2	5.38	0.93
	Av. weighed degree	5100.43	69.95	0.93
	Av. path length	1.87	3.87	6.41
	Diameter	4	8	37
	Density	0.19	0.04	0
	Modularity	0.89	0.92	1
	Number of communities	31	33	637
	Weakly connected components	13	16	637
	Strongly connected components	15	29	9241

that every student would have read all the messages posted by his or her colleagues (including their own posts). Values lower than those indicate that: (1) there are students who have not posted any message; (2) there are students who are not reading other students' posts; or (3) there are students who are posting but their messages are not being read by their peers. Instructors may use this information to further inspect if any of these three scenarios is happening (e.g., looking for low values of node in- or out-degree). Course 1 should have both general and intra-group social interaction in forums, and the data confirms that there is high reading activity among students; in other words, most of the students are reading each other's messages, but not all. Further inspection of data shows that one student is not reading any messages and other student's posts have only been read by five of his peers. In course 2, the values are much lower. However, it must be noticed that most of the social exchanges in course 2 are focused on the group forums, and that each student has access only to the general forums and their own group forum. That is, unless a student posts a message to the general forum, his or her in-degree will not be higher than the number of group members plus the number of teachers; conversely, the maximum out-degree value will be equal to the number of group members plus the number of different students posting to the general forums. In these cases, modularity can be a more interesting parameter to observe because it gives information about how strong are the links within a given component (connected subgraph). The modularity in teamwork-intensive courses should be close to 1.

Parameters of the Replies graph show information about active interaction between students. In this graph, some of the overall network values may provide little information besides general activity. However, results of individual centrality metrics in this graph are critical for detection of relevant agents in the learning process (see, for example, Fig. 11.8). Besides this, it is also important to observe modularity values, as well as the number of communities, weakly and strong connected components to observe whether groups are cohesive and whether there may be disconnected students who are not actively engaging in the course.

Finally, parameters in the Messages graph offer relevant information about how active are the threads in a course. First, the number of nodes indicates the total number of messages, and the number of communities/weakly connected components correspond to the number of different threads—because different threads have no posts in common. Then, the average degree or weighed degree represents the number of initial posts that have been left unanswered. Interestingly, in course 1, 4 out of 5 initial posts received some answer; upon inspection in the data laboratory, most of the unanswered posts were tidbits of information left by the teacher. Course 2 had 125 (6.89 %) posts with no replies, 99 of which were posts to a forum used as an assignment repository. This information is useful for instructors to check whether there are unsolved questions in courses with high forum activity—and it can be further enhanced if they consider information about message timestamp to distinguish between older and newer posts.

Conclusion

Following higher education students' information consumption and learning habits, and in order to profit from the advantages offered by information technologies (wider audience reach, ubiquitous access or lower costs, among others), current educational trends are leaning toward ICT-based and ICT-supported learning methods and approaches.

In online learning, the use of information technologies is intensive, and the instructional methods tend to focus on empowering self-directed and social learning. As the complexity of courses and the number of students enrolled in a course increase, tracking students' progress becomes a titanic task for instructors. The data stored in formal learning environments' databases contain valuable information that can make instructor's job much easier. However, these data are available in raw format, and further processing is required in order to provide meaningful and actionable data about the courses and the social dynamics that are associated to them.

Learning analytics is a new discipline that covers the collection, analysis, and reporting of these data to improve the learning process. Within this discipline, some approaches have specifically focused on the analysis of the social interactions occurring in ICT-mediated learning using SNA techniques, in what has been named social learning analytics. While social learning analytics is a broad term that supports different perspectives, one of the main concerns of researchers and practitioners has been how to embed social learning analytics features in existing formal ICT-based learning environments.

Our approach to social learning analytics in this chapter has been different: following the idea that tools that are built to solve specific problems are more suitable to address them, we point to general purpose SNA tools as a better alternative for social learning analytics. The main problem for the use of these tools in ICT-supported educational contexts is that LMS log databases are not built with social learning relationships in mind.

Therefore, this chapter introduced GraphFES, a web service and application that extracts information about forum activity from Moodle and builds GEXF files that represent a graph of the resulting passive and active user interaction networks, as well as a graph of the relations among the messages exchanged in the course forums. These files can be later processed using general purpose SNA tools (e.g., Gephi) for social learning analytics.

Along the chapter, we described the functionality, design, and implementation of GraphFES, and we showed and illustrated with data from real courses some of the possibilities of Gephi for social learning analytics. The chapter did not aim, though, to fully explore the features and capabilities of Gephi as social learning analytics tool, but rather to show the potential of SNA tools for in-depth social learning analytics. Despite some progress in the discipline, social learning analytics is still mostly a blank canvas on which researchers are beginning to create a new type of paintings. Our effort in this chapter and the development of GraphFES would be just the equivalent of providing sketching pencils that are appropriate for these new drawing techniques.

Nonetheless, the development of such techniques will still require further research on the different topics that fall under the term social learning analytics. In our opinion, there are three main prospective lines that would contribute to radical improvement of social learning analytics.

First, the differences in course characteristics and pedagogical approaches make it very difficult to find general rules of application for successful SNA. Social learning analytics is a new discipline, and deeper investigation on relevant indicators and optimal SNA metrics' values for each type of course is encouraged. Even though the results from our case studies are not easily generalizable, throughout section "Case Studies" we have given some guidelines about how interpret the results of the analysis in Gephi of data delivered by GraphFES in two different contexts. We strongly believe that visualization of Moodle data in Gephi may help teachers to easily and rapidly detect disconnected and engaged students, especially in intensive teamwork or project-based learning, and that SNA metrics may help refine visual results with a higher degree of detail and facilitate complementary analysis for researchers. In this sense, GraphFES is just a tool that facilitates social learning analytics but, from a wider perspective, successful implementation of social network analytics across an institution requires not to consider social learning analytics as a convenient tool, but rather a part of an integral learning analytics strategy, taking into account the different learning methods and objectives.

Second, the potential of social learning analytics for the improvement of teaching and learning is enormous. However, the use of SNA tools is not simple or easy without some degree of training and understanding of the different concepts involved. There are three ways to circumvent this barrier: giving adequate training in SNA techniques to instructors and teachers, lowering entry barriers by providing teachers with basic training and a handful of useful predefined filters, and having a layer of advanced SNA users that can act as consulting advisors of teachers and help them get and understand the relevant information from the analysis of their courses. Considering the three options, the first one may be time- and cost-consuming, and thus institutions should consider a choice between the other two in their learning analytics strategy.

Third, the current initial version of GraphFES offers enough functionality for SNA of educational data, but further improvement of the tool is still possible. Future development of the tool is required to expand its capabilities, such as: collection of other information that might be of interest for researchers and practitioners, and integration of that information as node or edge attributes; transformation of temporal data—currently, only the timestamp of posted messages is collected—to build dynamic graphs that allow observation of changes in course dynamics—Gephi has the ability to build such timelines—; improved aggregation of semantic data about messages and their contexts to facilitate semantic analysis and discourse analytics; and finally, new versions of the tool should also take advantage of any new functions that may be added to Moodle's web service layer.

References

Amo Filvà, D., García-Peñalvo, F. J., & Alier Forment, M. (2014). Social network analysis approaches for social learning support. In *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM '14)* (pp. 269–274). New York: ACM.

Bandura, A. (1971). *Social learning theory*. New York: General Learning Press.

Barkley, E. F., Cross, K. P., & Major, C. H. (2005). *Collaborative learning techniques: A handbook for college faculty*. San Francisco, CA: Wiley.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the Third International ICWSM Conference* (pp. 361–362).

Beaudoin, M. F. (2002). Learning or lurking? Tracking the "invisible" online student. *The Internet and Higher Education, 5*(2), 147–155.

Berger, P. L., & Luckman, T. (1967). *The social construction of reality: A treatise in the sociology of knowledge*. Harmondsworth: Penguin.

Buckingham-Shum, S., & Ferguson, R. (2012). Social learning analytics. *Journal of Educational Technology & Society, 15*(3), 3–26.

De Laat, M., & Prinsen, F. (2014). Social learning analytics: Navigating the changing settings of higher education. *Research & Practice in Assessment, 9*, 51–60.

Edutechnica. (2015). *LMS Data–Spring 2015 Updates*. Retrieved from http://edutechnica.com/2015/03/08/lms-data-spring-2015-updates.

Fidalgo-Blanco, Á., Sein-Echaluce, M. L., García-Peñalvo, F. J., & Conde, M. Á. (2015). Using learning analytics to improve teamwork assessment. *Computers in Human Behavior, 47*, 149–156.

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks, 1*(3), 215–239.
Gephi. (2015). *Supported Graph Formats*. Retrieved from http://gephi.github.io/users/supported-graph-formats.
GEXF Working Group. (2015). *GEXF File Format*. Retrieved from http://gexf.net/format.
Hernández-García, Á. (2014). Usare Gephi per visualizzare la partecipazione nei corsi online: Un approccio di social learning analytics. *Tecnologie Didattiche, 22*(3), 148–156.
Hernández-García, Á., & Conde, M. Á. (2014). Dealing with complexity: Educational data and tools for learning analytics. In *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM '14)* (pp. 263–268). New York: ACM.
Hernández-García, Á., González-González, I., Jiménez-Zarco, A. I., & Chaparro-Peláez, J. (2015). Applying social learning analytics to message boards in online distance learning: A case study. *Computers in Human Behavior, 47*, 68–80.
Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating collaborative knowledge building. *Cognition and Instruction, 26*, 48–94.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
Leris, D., Fidalgo, Á., & Sein-Echaluce, M. L. (2014). A comprehensive training model of the teamwork competence. *International Journal of Learning and Intellectual Capital, 11*, 1–19.
Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 31–40.
Macfadyen, L., & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Journal of Educational Technology & Society, 15*(3), 149–163.
Moodle. (2015). *Logging 2*. Retrieved from https://docs.moodle.org/dev/Logging_2.
Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge building discourse explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research and Development, 60*(5), 903–921.
Reffay, C., & Chanier, T. (2003). How Social Network Analysis can help to Measure Cohesion in Collaborative Distance-Learning. In B. Wasson, S. Ludvigsen, & U. Hoppe (Eds.), *Designing for Change in Networked Learning Environments: Proceedings of the International Conference on Computer Support for Collaborative Learning 2003* (pp. 343–352). Dordrecht: Springer Netherlands.
Wise, A. F., & Hausknecht, S. N. (2013). Learning analytics for online discussions: A pedagogical model for intervention with embedded and extracted analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)* (pp. 48–56).
Zhao, K., & Chan, C. K. K. (2014). Fostering collective and individual learning through knowledge building. *International Journal of Computer-Supported Collaborative Learning, 9*(1), 63–95.