Social Network Analysis on Educational Data Set in RDF Format

Bogdan Drăgulescu, Marian Bucos and Radu Vasiu

Faculty of Electronics and Telecommunications, Politehnica University Timişoara, Timişoara, România

The increased usage of information technologies in educational tasks resulted in high volume of data, exploited to build analytical systems that can provide practical insight in the learning process. In this paper, we propose a method of running social network analysis on multiple data sources (academic years, communication tools). To achieve this, the collected data that describe social interactions were converted into a common format by employing a prior developed semantic web educational ontology. Using a mapping language the relational data set was linked to the appropriate concepts defined in the ontology and then it was exported in RDF format. The means for SPARQL access was also provided. Subsequently, query patterns were defined for different social interactions in the educational platform. To prove the feasibility of this approach, Gephi tool set was used to run SNA (Social Network Analysis) on data obtained with the SPARQL queries. The added value of this research lies in the potential of this method to simplify running social network analysis on multiple data sets, on a specific course or the entire academic year, by simply modifying the query pattern.

Keywords: social network analysis, learning analytics, semantic web, educational ontology

1. Introduction

Internet technologies have large-scale penetration thus pushing the development of electronic versions for classic services (mail, publications, commerce, etc.). Matter-of-course, this process has reached the educational field, where electronic systems that can offer support in the pedagogical process have emerged. They use the Internet as a communication medium.

Organizations that offer educational programs for distance learning accelerated the development of educational platforms. For these, using electronic educational systems it is a logical step in offering a better service. On the other side, companies showed an increased interest for e-Learning technologies, in order to reduce costs with staff training. Universities that offer traditional face-to-face educational programs tend to use e-Learning to improve access to course materials, to offer personalized support for students to decrease dropout rates, and to promote themselves on international markets. (Naidu, 2006)

Custom software tools are needed for electronic support in the pedagogical process. These can be integrated into a platform to offer a complete solution, or can be independent tools used by the tutor in the educational process. For a big organization that offers a large number of courses, the first solution is more suited. In this context, the Learning Management System's (LMS) have emerged (Sclater, 2008). There are open source solutions, a popular one being Moodle, and also commercial counterparts, such as the powerful Blackboard platform. Each one of these approaches has advantages and disadvantages, so the best choice for an organization is determined by its needs.

At our university, based on the experience from a previous research project (ViCaDiS – Virtual Campus for Digital Students), the Moodle platform was chosen to implement the University's e-Learning platform. At first, the platform was used for the distance learning program, and then extended for bachelor and master programs. At the current time the platform holds more than 1000 courses and has 11 439 users. In general, it is considered that the educational process

evaluation is time and resource consuming and is presently lacking the right tools to automate these tasks.

Also, in recent years a new paradigm called Massive Open Online Courses emerged, in which the course developers offer free access to educational resources. This type of courses integrates social networks, online resources and are facilitated by experts in the course domain. By providing the courses over the Internet and free of charge, they tend to have a large pool of users (McAuley et al., 2010). In this context, such tools, that automate analytics tasks regarding the educational process, become a necessity.

This increased usage of information technologies in the educational process has led to high volume of data exploited to build analytical systems that can return reports, alerts, models, recommendations, and predictions, which are relevant for students, tutors or system administrators. In a report regarding the evolution of data analysis in educational platforms (Cooper, 2012), Adam Cooper described research fields from which Learning Analytics borrows methods. Some examples were: statistics to test a hypothesis, business intelligence to process and present the data, machine learning and data mining for identifying patterns in the data, social network analysis to measure students' involvement, and data visualization.

George Siemens, in an article from 2012 (Siemens, 2012), described his vision for Learning Analytics research domain development, stressing the interdisciplinarity of the field. The author argued the high potential of the field to influence pedagogical models and to offer a better understanding of the causes that make a model succeed or fail. To take advantage of this potential, new methods and tools need to be developed. Additionally, privacy issues regarding user anonymity must be addressed.

This study proposes a method of running social network analysis using multiple data sources. The major challenges addressed in this paper are: obtaining a coherent data structure and defining a simple means of extracting the needed information for different scenarios. To achieve this, the use of a prior developed educational ontology and semantic web technologies is proposed.

The paper is organized as follows: Section 2 shows background about running SNA on educational data and some related works. Section 3 presents our approach to running SNA on multiple data sources. Section 4 describes the process of building a coherent data set by using an educational ontology. Section 5 shows the defined SPARQL patterns to extract data for SNA. In Section 6 there is a discussion of this work, and finally Section 7 summarizes the main conclusions.

2. Background

One technique that can be applied in Learning Analytics is Social Network Analysis (SNA). It offers a method to focus on social interaction data and not on the characteristics of the individual. Thus the unit of analysis is the interaction that occurs among members of the network (Laat et al., 2007).

Educational researchers are taking advantage of the growing application of computer assisted education and use the data collected through such systems to test the impact of social interaction in the pedagogical process.

In an early article (Laat et al., 2007), the authors suggest using SNA to visualize participants gain or loss of interest during a collaborative task. They argued that this technique adds valuable analytical information for understanding the processes occurring in learning communities.

In another paper, Dawson (Dawson, 2008) has employed SNA techniques to explore the relationship between the student position in a classroom social network and their sense of community. The author has found that closeness and degree are positive predictors of an individual's sense of community.

Another study aimed to investigate the potential relationship among student admission criteria, exam performance and engagement in the learning community. They used Social Network Analysis to evaluate the effectiveness of student admission criteria (Dawson et al., 2011).

Gašević et al. presented an empirical study (Gašević et al., 2013), where the authors analyzed 10 years of data in a master's degree program offered through distance education, and proved that students social capital is positively associated with their academic performance.

Other authors (Ferguson and Shum, 2012; Shum and Ferguson, 2012) present the advantages of social network analysis, and define the technique as Social Learning Analytics. They found that knowledge acquisition is influenced by social interactions. Data used for analysis are generated by direct interactions in communication tools, grading systems and activities that influence the actions of other participants. Five categories are described: network analytics, discourse analytics, content analytics, disposition analytics and context analytics.

2.1. SNA Tools

Evaluating students' interactions in Learning Management Systems is difficult, resource demanding and warrants expertise. To use the above findings in the pedagogical process by the actors involved (tutors, students and administrators), custom tools need to be built.

A network learning analytics concept tool was described (Retalis et al., 2006), stipulating also the necessity of such an approach in the Moodle LMS. The CosyLMSAnalytics tool allows the evaluator to easily track students' behavior, gather feedback in an automatic way and present the information in a comprehensive fashion. The software can export the data in adequate format for an external tool for SNA.

A social network analysis software program that can be used for educational purposes is UCINET. It enables users to analyze network routines and general statistics of multivariate analyses (Borgatti et al., 2002). The software was designed for graph and network analysis, but can also be used for SNA in education. One example of such a software is GEPHI, an open source tool that uses a 3D render engine to display large networks in real-time, to speed up the exploration, allow spatializing, filtering, navigating, manipulating and clustering (Bastian et al., 2009). Another open source graph visualization software that can be used for SNA is GRAPHVIZ (Ellson et al., 2004) .

To use the above tools, the users have to obtain the data, preprocess and convert it into the suitable format, load in the data and then learn to use the tools. For the common education practitioners this can be an overwhelming process that requires some expertise.

SNAPP is a tool that infers networks from forum structures, designed as a client-side bookmarklet, requiring only a web browser, with no database access or server-side scripting needed (Dawson et al., 2010). The benefits of this tool are: simple installation and usage, multiple LMS support, cross browser and platform support, real-time forum data analysis, in-line visualization and social network metrics, multiforum support and simplified SNA concepts for users that are not SNA experts. Though this tool is user-friendly, it cannot use data that the instructor has no browser access to, and in this way it cannot analyze the data in a multi-course or archived courses environment.

2.2. Related Work

One approach in using semantic web technologies to integrate data from multiple discussions sites is presented in the paper of Hover and Muhlhauser (Hover and Muhlhauser, 2013). The authors used enterprise integration patterns for seamlessly integrating the discussion forums from Moodle and the social network Facebook. Additionally, they have shown how discussions can be easily integrated in lecture recordings on data level.

Cantador and Castells in their paper (Cantador and Castells, 2006) presented an approach to automatic identification of social networks. They used an ontology to take into account the semantic preferences of several users, clustering the ontology concept space and obtaining common topics of interest. Using this clustering strategy, they ran experiments to test the feasibility of this method for collaborative filtering models.

An ontology model for social learning analysis was also proposed in other papers (Yusof et al., 2011; Yusof and Mansur, 2013). The authors designed the ontology to capture the information contained by their university's educational platform based on Moodle. All the communication tools and related data are modeled by using custom defined classes and properties, so the reuse of well established ontologies was not taken into account, although this is a desirable practice for large-scale adoption of semantic web technologies (Simperl, 2009).

3. Approach

In this paper we propose a method for running social network analysis on converted relevant educational data, collected by the LMS, in RDF format. With this approach, we aim to address the problem of using multiple data sources (archived versions of the educational platform, external data sources) and we provide a solution for accessing the relevant data.

According to Chatti et al. (Chatti et al., 2012), a learning analytics process can be summarized in three major steps: collecting and pre-processing data, analysis and decision, post-processing. The first step involves data collection, cleaning the irrelevant data, data integration, and modelling the data for future steps. In our approach we use a semantic web ontology developed in our previous research (Bucos et al., 2010; Dragulescu, 2012) for republishing educational data in a unified model to easily merge data from different sources. In the next section we present building of the data set from our university's educational platform, which data is relevant to social network analysis. In the second step the data set is analyzed and patterns are detected using Learning Analytics techniques. For our data model we adapted and tested these techniques using Semantic Web technologies. The last step requires evaluation and optimization, and is fundamental to the continuous improvement of the analytical process.

4. Building the Data Set

As the data source for this experiment we used our university's educational platform based on Moodle. At the moment, the methodology to prepare the platform for a new academic year involves creating an archive version for administrative reasons, upgrading to the latest Moodle version and in-house developed modules, resetting the courses and enrolling the students. This implies modification of the data storage structure, making multi-instance data analysis a complicated task. This is one consideration for which we proposed a unified structure to store the educational data for learning analytics.

Table 1 shows evolution of the educational platform regarding Moodle version used, number of the courses implemented, number of student accounts, number of teacher accounts and instances of communication tools. At first, the platform was used for the distance learning program. Increase of the courses available through the platform is mainly due to the university policy to extend the use of the platform for all master programs. The increase of users correlates with the increase of courses offered through the platform. The main communication tools used were forum and OU blog. What is not present in the table, and might perhaps contain useful information for social network analysis, is the communication among users outside of a course using the messaging module.

The ontology proposed by Dragulescu in his PHD thesis (Dragulescu, 2012), was used to convert data from multiple instances of the educational platform. The ontology is composed of four sub-parts specific to the modeled data. The Person sub-ontology extends the FOAF vocabulary and standardizes the concepts regarding the users. Some useful concepts for the social network analysis are the classes *edu:Student* and

	2010 – 2011	2011 – 2012	2012 – 2013	2013 – 2014	2014 – 2015*
Moodle version	1.9.9+	2.1.x	2.2.x	2.5.x	2.7.x
No of Courses	577	773	878	1267	1285
No of Students	2568	3694	3664	6342	6778
No of Tutors	262	378	427	575	573
Chat	17	8	42	52	48
Forum	633	812	955	1389	1450
Ou Blog	1299	1282	1308	1330	1329
Wiki	174	140	182	181	197

*only the first semester

Table 1. Instances of communication tools over the last 5 academic years.

edu:Tutor that define the role of the users in the educational platforms; they are subclasses of *foaf:Person*. Properties *edu:hasStudent* and *edu:hasTutor* are used to define tutor-student and student-teacher relationships. To represent the same role relationships, student-student or teacher-teacher, the *edu:hasColleague* is used. All the above properties are sub-properties of *foaf:knows*.

Course sub-ontology formalizes the concept of a course by defining the class *edu:Course*. Besides this concept, the classes *edu:Curriculum* and *edu:SectionOfCurriculum*, with the corresponding properties *edu:belongsToSectionOfCurriculum* and *edu:hasSectionOfCurriculum*, are of interest for SNA because they can be used to filter the data for a specific educational program. If the analysis is on a specific part of a course, the class *edu:SectionOfCourse* and the corresponding property *edu:hasSectionOfCourse* can be used.

For this application the most important part of the ontology is the *Activity* part. It formalizes the concepts needed to describe all the tools available in the educational platform, from assignments to communication tools. The latter is of the greatest interest and is formalized with the class *edu:CommunicationTools*, a sub-class of *edu:Activity*. To model interaction in the communication tools, the SIOC ontology was employed. Particularly useful are the classes *sioc:Container*, with the sub-class *sioc:Forum*, and corresponding properties *sioc:has_container*, *sioc:container_of*, *sioc:has_parent* and *sioc:creator_of*, used to describe the structure of the discussion medium. To describe one message in a discussion we employed the class *sioc:Item* or sub-class *sioc:Post*. Properties *sioc:reply_of* and *:has_reply* keep the relationship between posts. The properties *sioc:has_creator*, *sioc:created_at*, *sioc:subject* and *sioc:content* are used to describe the message.

Finally, the fourth sub-ontology, called *Evaluation*, has at the moment no application for social network analysis.

Figure 1 depicts concepts from the educational ontology used to model the data for SNA analysis. It contains the classes with the subclass relationships, the properties with range and domain definition. The SIOC ontology classes and properties are not depicted in this model.
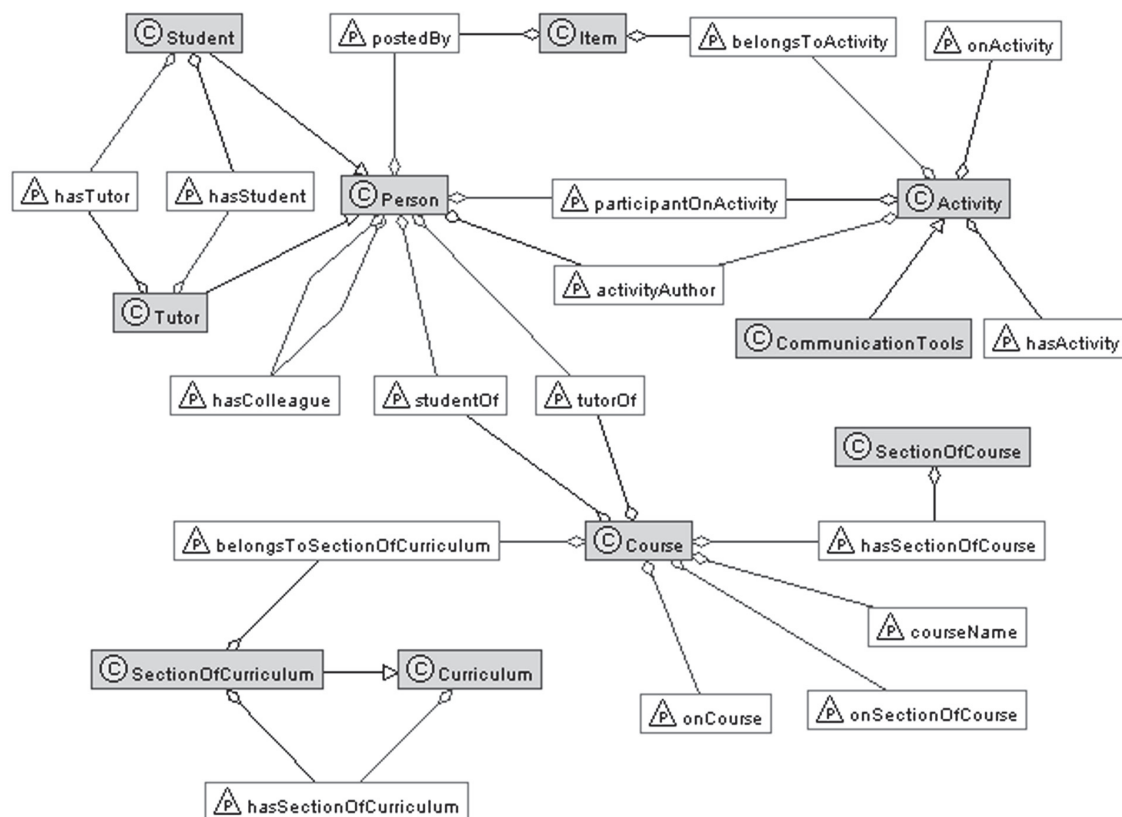

Figure 1. Fragment from the educational ontology. Classes and properties used for SNA.

After we identified the information of interest for our experiment and the ontology that described necessary concepts, the next step was to convert the data from SQL format to RDF. To accomplish this task, we employed the D2RQ platform. This software defines a way to access relational data by using a declarative language to map the ontological concepts to SQL tables and attributes (Bizer and Seaborne, 2004). Because the versions of Moodle differed from one academic year to the next, and some changes were made in each subsequent upgrade to the database, we created a mapping file and adapted it for each instance.

To access the data in RDF format, the D2RQ platform provides a stand-alone server and a JAVA war application. We chose the latter and deployed it on a Tomcat server. The data set was made available throughout a SPARQL endpoint and the integrated SNORQL query explorer interface was used for testing purposes. In this approach, the data is stored in the default relational model The tool converted the SPARQL call to an SQL query, transformed the returned data to RDF format and then sent it to the client. This data query model ensures that the data are always up to date.

Because we had five instances of the educational platform with five corresponding mapping files, this implies that we had to have five deployments of D2RQ application. This was not the case. For the archived instances of the platform the data are static. Therefore, we exported the data in RDF format for the archived versions, and loaded it in a triples store. For this task we employed the OpenRDF SESAME framework version 2.7.9, deployed in the same Tomcat environment. The RDF dump files were imported

in RDFS native store for persistent storage, each file in their specific context to keep track of data provenance. Thus, the data was available in an RDF format through two endpoints, one from the D2RQ server that mapped the live platform database, and one from the SESAME server that stored the RDF data from the archived instances. Figure 2 depicts the production flow of the RDF data set.

Following the method described above, approximately 1.8 million triples were made available in RDF format for the academic year 2014-2015, 1.7 million for 2013-2014, 0.9 million for 2012-2013, 0.8 million for 2011-2012 and 0.5 million for 2010-2011. The resulting number of instances for classes and number of triples for properties used to describe the communication tools over the last five academic years, are shown in Table 2. For the current academic year the data was presented only for the first semester.

5. Obtaining Data for SNA with SPARQL

To extract the data for social network analysis from the new RDF format, the appropriate query patterns needed to be defined. The data set contained social interaction in three communication tools from the educational platform: forum, blog and personal messages. Because the data from these communication tools was converted using different concepts and had a different hierarchical structure, a general query pattern was defined for each tool.

Below we represented a pattern for extraction of social interactions in all the forums from a specified course, for a a limited time frame in
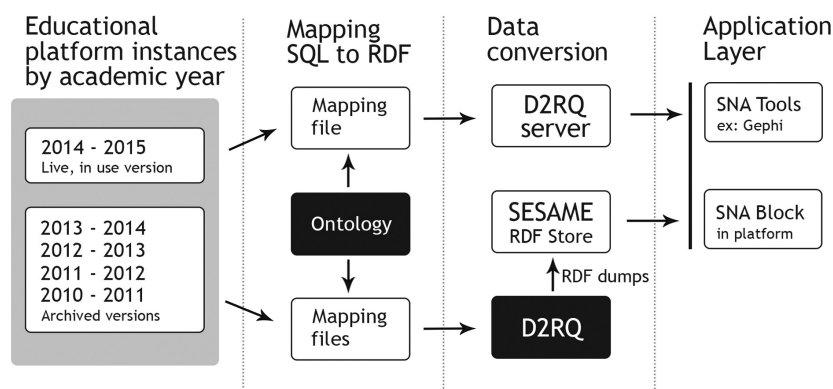

Figure 2. RDF conversion process.

	2010–2011	2011–2012	2012–2013	2013–2014	2014–2015*	Type
sioc:Post	2321	1575	2019	2940	1689	class
sioc:Container	1299	1282	1308	1330	1329	class
sioc:Thread	681	581	847	1261	1020	class
sioc:Forum	633	812	955	1389	1450	class
sioc:Item	1498	3465	7525	16633	19715	class
sioc:reply_of	630	315	413	711	376	property
sioc:has_creator	3819	5040	9544	19573	21406	property
sioc:created_at	2321	1575	2019	2940	1689	property
sioc:title	1037	707	788	1053	298	property
sioc:content	3819	3510	8011	18039	19872	property
sioc:has_container	3695	4952	9452	19397	21371	property
sioc:has_parent	681	581	847	1261	1020	property
sioc:addresed_to	1498	3465	7525	16633	19717	property
sioc:subject	1199	820	1180	1808	1373	property
edu:hasActivity	3070	3355	3789	4682	4802	property
edu:activityName	3751	3936	4636	5943	5822	property
dc:created	2179	4046	8372	17894	20737	property

*only the first semester

Table 2. Number of instances for classes and number of triples for properties.

an academic year. The pattern was expressed using SPARQL algebra. Only the replies to a forum post were considered social interactions, and thus one condition was that the two posts needed to be linked to the property *sioc:reply_of.* The posts needed to be forum posts. This was accomplished by specifying that the container of the first post was an instance of the class *sioc:Thread*, used only for forums. To obtain only the interactions from a specific course the property *edu:hasActivity* was used to link the parent forum of the discussion to a specific URI course. The time limit was imposed by first obtaining the published time of the first post and then by adding a condition in the filter section of the query. Another filter condition was that the author of the first post and the author of the second post must be different. In this way, the self replies were eliminated.

Forum query pattern in SPARQL algebra

```
(prefix ((sioc: <http://rdfs.org/sioc/ns#>)
   (rdf: <http://www.w3.org/1999/02/22-rdf
      -syntax-ns#>)
   (edu: <http://elearning.upt.ro/ontology
      /edu#>))
(project (?person2 ?person1)
```

```
(graph <http://academic/year/context/uri>
 (filter (exprlist (!=?person1 ?person2)
      (>?time time_in_seconds))
  (bgp
   (triple ?post1 sioc:has_creator ?person1)
   (triple ?post1 sioc:has_container
      ?discussion)
   (triple ?discussion rdf:type sioc:Thread)
   (triple ?post2 sioc:reply_of ?post1)
   (triple ?post2 sioc:has_creator ?person2)
   (triple ?discussion sioc:has_parent ?forum)
   (triple <http://course/uri>edu:hasActivity
      ?forum)
   (triple ?post1 sioc:created_at ?time)
   )))))
```

Specifying the context in which data from a particular academic year was loaded limits the pattern to only extract data from that period. Based on the above conditions, the query returned all the forum interactions between two persons as URI pairs. The query can be adapted to obtain extra information needed for visualization. For example, the *foaf:lastName* property can be used to obtain the last name of the persons in order to be displayed as node labels.

For different scenarios, some of the above conditions can be eliminated or modified. By removing the course limiting triples, the query produces all the forum interactions in the specified academic year and time frame. This is useful for academic analytics, as by analyzing and visualizing the data for the whole system, reports are generated that can influence policies.

The second pattern, used to extract social interaction in all blogs from a specified course, is depicted below in SPARQL algebra. Only the comments to a blog post were interpreted as social interactions. This condition was implemented by requiring that the posts were linked with the property *sioc:reply_of*, same as for the forum pattern. Moreover, the first post must be an instance of the class *sioc:Container*, used in our implementation only for blogs. Similarly to the forum pattern, the property *edu:hasActivity* was used to link the blog to a specific URI course. The time for the first post was used to filter the results for a custom time interval. Also, the author of the first post and the author of the second post must differ. Finally, the context was used to limit the working data set to only a specific academic year. The statements from the forum query, regarding obtaining extra information for visualization purposes and modification of the pattern for different scenarios by removing conditions, remain valid for the blog query.

Blog query pattern in SPARQL algebra

(prefix ((sioc: <http://rdfs.org/sioc/ns#>)
   (rdf: <http://www.w3.org/1999/02/22-rdf
     -syntax-ns#>)
   (edu: <http://elearning.upt.ro/ontology
     /edu#>))
(project (?person2 ?person1)
 (graph <http://academic/year/context/uri>
  (filter (exprlist (!= ?person1 ?person2)
    (>?time time_in_seconds))
   (bgp
    (triple ?post1 sioc:has_creator ?person1)
    (triple ?post1 sioc:has_container ?topic)
    (triple ?topic rdf:type sioc:Container)
    (triple ?post2 sioc:reply_of ?post1)
    (triple ?post2 sioc:has_creator ?person2)
    (triple <http://course/uri>edu:hasActivity
      ?topic)
    (triple ?post1 sioc:created_at ?time)
    )))))

Lastly, for extracting social interaction among users outside a course, using the messaging module, the query pattern in SPARQL algebra is shown below. To obtain the person from which the interaction started, the property *sioc:has_creator* was used, and to obtain the person to whom it was addressed, the property *sioc:addressed_to* was employed. To limit the data for a specific time frame, a filter condition was set in place for message time and the context was specified for the desired academic year. There was no direct way to obtain only the messages that were in connection with a course. A useful approach is to add the social interaction from messages module that involve the participants in a given course, disregarding the fact that the messages may not address any subject of that course.

For all the above queries the context filter is not needed for the live version of the data set, because it is provided by the D2RQ server.

Messages query pattern in SPARQL algebra

(prefix ((sioc: <http://rdfs.org/sioc/ns#>)
   (dc: <http://purl.org/dc/terms/>)
   (rdf: <http://www.w3.org/1999/02/22-rdf
     -syntax-ns#>)
   (edu: <http://elearning.upt.ro/ontology
     /edu#>))
(project (?person1 ?person2)
 (leftjoin
  (graph <http://academic/year/context/uri>
   (filter (>?time time_in_seconds)
    (bgp
     (triple ?message1 sioc:has_creator
       ?person1)
     (triple ?message1 sioc:addresed_to
       ?person2)
     (triple ?message1 dc:created ?time)
    )))))

5.1. Case Study

To test the SPARQL queries, the Gephi open source software was employed. Gephi has been successfully used for internet link, semantic network case studies, and social network analysis (Bastian et al., 2009). The semantic web import plug-in was used to extract the needed data for SNA, using a built-in REST endpoint driver.
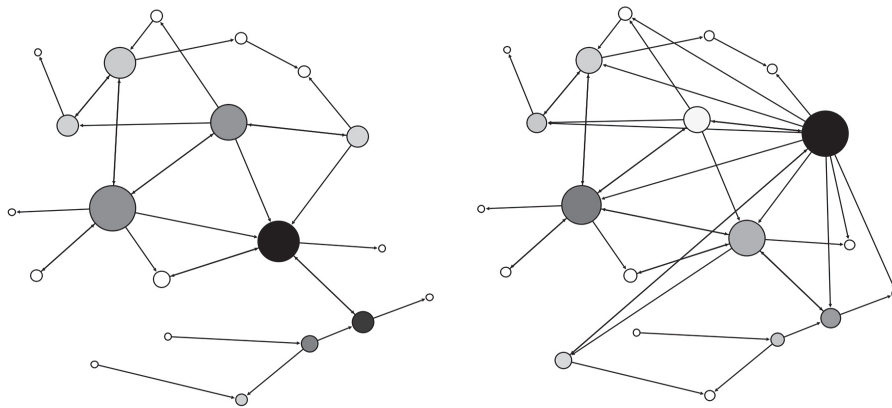
Figure 3. SNA on an example course. Only forum posts (left) versus forum and IM posts (right).

An example of a Gephi generated social network diagram, corresponding to a course from the current academic year, is shown in Figure 3. To visualize the benefits of using multiple data sources in a social network analysis we chose to run SNA in two different scenarios. On the left side only the forum post replies were used, versus the right side where the interactions from private messages were also added. The above forum and messages query patterns were adapted and used in the semantic import plug-in to extract the data in a graph. In both queries, the current academic year URI context identifier was specified. Also, a URI identifier for the studied course was used. Because in this particular case the analysis was run on all the data collected in a course, the time filter was not specified.

The nodes were not moved from one evaluation to the other. Degree centrality is shown through the size of the node. The person represented by the biggest node has the most social ties in the social network of this course. The darker filled nodes have a high value of betweenness centrality. This metric quantifies the shortest paths between any two nodes that go through that given node. People with higher degree centrality are more likely to have higher access to social resources, which can enable them to enhance their performance (Sparrowe et al., 2001). High betweenness centrality is associated with brokerage positions: the individuals in these positions can control information flow and have potential to produce more innovative results (Burt, 2004). Comparing the two social network diagrams, one can observe the shift in the nodes with the highest betweenness centrality and degree centrality. From this an educator

can draw different conclusions and makes more informed decisions.

Though this example is run on a low scale (only the data collected in a course), the same principles can be applied in different cases. The same queries can be run without the course filter to obtain the data for social interactions from the entire academic year. Alternatively, by removing the context, SNA can be run on all available data.

6. Discussions

As the process of communication develops, a person positioned in the social graph that has direct contact with many others should begin to see himself and be seen by others as a major source of information. At the opposite extreme, the occupant of a position with low degree is perceived as peripheral, and is isolated from direct involvement with most of the others on the network (Freeman, 1979). On the other hand, according to Lockyer et al. (Lockyer et al., 2013), for interpreting the data that learners' environments generate, it is important to combine learning analytics with learning design. For example, the learning design and intent of a forum influences the interpretation of analytics visualization. In a question and answer forum the instructor tends to be the central actor, conversely, in a discussion forum the facilitating students tend to be in the central positions. Thus, the interpretation of the social learning analysis for a particular course is best performed by the corresponding tutor in accordance with the learning design employed.

Some of the proved scenarios that offer an insight into the educational process, and involve cross-class social networks, are: investigating the relationship between student academic performance and social ties (Gašević et al., 2013), community development in distance education programs (Haythornthwaite et al., 2000), or evaluating learning communities (Lenning and Ebbers, 1999). These are some examples where a trained practitioner with the role of manager (anyone responsible for the organization and administration of teachers (Clow, 2012)), can interpret the social network analysis of a cross-class multiple academic year data set. Based on that analysis, the managers can take necessary actions to promote initial bonding, monitoring and supporting social interactions, and providing multiple means of communication, if needed.

The method for social learning analysis proposed in this paper is suited for both levels of access described above. The synthesized steps are: identifying collected data relevant for the specific analytic task; deciding which semantic web ontology is best suited; mapping the data to the appropriate concepts from educational ontology; exporting the data in RDF format and loading it into a RDF store using different context for each academic year; designing the query patterns; using a social network analysis tool set that can obtain data from RDF endpoints. For the current academic year the best approach is not to export the data, but to provide means to return it in RDF format on request. In this way, access to the latest data is ensured.

This approach ensures that the data is available in the same format for all academic years, regarding the changes in the educational platform data structure. Thus there is no need to change data extraction and processing pattern for the social network analysis on every platform upgrade.

By converting educational data into a coherent data structure, extraction of the needed information for different levels of access or from multiple academic years can be done by only modifying the query pattern. Using the course identification URI with the appropriate pattern, the data returned is usable for analytics done by the tutor of that course. From the manager point of view, he can employ the context, time, or any combination of patterns to obtain the data relevant for his analysis.

The manager has the right to run social network analysis on all the data collected by the educational platform, conversely, the tutor's access has to be limited to the courses that he teaches. One approach is to limit access to RDF data based on the user privileges defined in the educational platform by mapping and converting the user privileges to RDF. The privileges in RDF format can be used with an access control specification language to enforce the access rules (Costabello et al., 2012; Flouris et al., 2010). For this method to work, validating the user identity has to be done using a login procedure. This is not currently available in the Gephi tool set, so an extra plug-in needs to be developed. The second approach is to rely on the access control module from the educational platform to confirm the user and check the rights of the tutor. For this, a module for the educational platform needs to be developed that can query the RDF endpoints and provide visualization of the social interaction graph.

On the other hand, interpreting the social interaction graph needs prior knowledge, and some skills are required to run the analysis using a specialized tool like Gephi. The manager is more suited to use such a tool. For the teacher, however, the analysis must be run in an automatic way and some interpretation of the social graph must be presented. This suggests that the best approach to provide access to social network analysis to the tutor is to build a module for the educational platform. In this way, one addresses the concerns regarding access control and the skills required to run SNA.

Controlling access to the RDF endpoints can be done by using the built-in Tomcat authentication or restricting access to specific IP lists. This is a mandatory step if confidential information is available through the endpoints.

The proposed approach was tested on our University educational platform based on Moodle, but it is not limited to this specific LMS. It can be used in any educational software that collects SNA relevant data. The key difference in adapting this approach for a different educational software is to correctly identify the relevant data and map it to the appropriate concepts from the ontology. The exporting, storing, querying and running SNA steps remain the same. Also, the use of the above defined ontology is not mandatory. Any predesigned or

custom ontologies can be used with some considerations. First, if the scope is to run SNA on data from multiple sources the same ontology must be used to convert the data. And second, if the data is to be shared with different parties, the best approach is to use a well-known mature ontology like SIOC.

With the addressability of this approach to any educational software that collects relevant data for SNA, we believe that this method can simplify the social network analysis on multiple educational data sets. This is achieved by providing a means for coherent data structure, storage and query.

We intend to employ this method to build a module for SNA in our educational platform. One possible benefit of using RDF data that is worth investigating is the possibility of adding multiple data sources from linked data, and study what implication arises. One case study is to add the social interaction from social networks, and evaluate the impact of this approach on the social graph. A natural future direction of research is to investigate if using RDF data for analysis is feasible for other learning analytics scenarios too.

7. Conclusions

This paper presents a method of running social network analysis by using RDF data sets. A prior developed educational ontology is used to model the necessary data from our learning management system based on Moodle. Data collected from five academic years were mapped to the concepts defined in the ontology by employing the mapping language introduced in the D2RQ framework. Two approaches were used to make the information available in RDF format: for static data the information was exported in RDF format and loaded in a RDF store, for live data on request RDF conversion was employed to ensure access to the latest version. SPARQL query patterns were defined to obtain the social interaction. Gephi tool set was used to prove the feasibility of this approach to run social network analytics for pedagogical purposes.

The significance of this research lies in the potential of this method to simplify running social network analysis on multiple data sets. By providing a unified data structure the needed information can be obtained by running the appropriate SPARQL query pattern. In this way the same data model can be used to obtain the social interactions from archive courses, multiple communication tools, from a specific course or the entire academic year, by simply modifying the query. Further research is needed to test the possibility of extending this approach to other learning analytics scenarios.

Acknowledgments

References

[1] M. BASTIAN, S. HEYMANN, M. JACOMY, Gephi: an open source software for exploring and manipulating networks. Presented in the *Proceedings of the 3rd AAAI Conference on Weblogs and Social Media*, (2009) San Jose, CA.

[2] C. BIZER, A. SEABORNE, D2RQ-treating non-RDF databases as virtual RDF graphs. Presented in the *Proceedings of the 3rd International Semantic Web Conference*, (2004) Hiroshima, Japan.

[3] S. BORGATTI, M. EVERETT, L. FREEMAN, Ucinet for Windows: Software for Social Network Analysis, *Analytic Technologies*, 2002.

[4] M. BUCOS, B. DRAGULESCU, M. VELTAN, Designing a semantic web ontology for E-learning in higher education. Presented at the *9th International Symposium on Electronics and Telecommunications*, (2010) Timisoara, Romania.

[5] R. S. BURT, Structural holes and good ideas, *American journal of sociology*, 110 (2004), 349–399.

[6] I. CANTADOR, P. CASTELLS, Multilayered semantic social network modeling by ontology-based user profiles clustering: Application to collaborative filtering, *Presented at the Managing Knowledge in a World of Networks*, (2006) Podebrady, Czech Republic.

[7] M. A. CHATTI, A. L. DYCKHOFF, U. SCHROEDER, H. THÜS, A reference model for learning analytics, *International Journal of Technology Enhanced Learning*, 4 (2012), 318–331.

[8] D. CLOW, The learning analytics cycle: closing the loop effectively, Presented at the *2nd International Conference on Learning Analytics and Knowledge*, (2012) Vancouver, BC, Canada.

[9] A. COOPER, A brief history of analytics, *Analytics Series 1*, 2012.

[10] L. COSTABELLO, S. VILLATA, F. GANDON, Context-Aware Access Control for RDF Graph Stores, Presented at the *biennial European Conference on Artificial Intelligence*, (2012) Montpellier, France.

[11] S. DAWSON, A Study of the Relationship between Student Social Networks and Sense of Community. *Educational Technology & Society*, 11 (2008), 224–238.

[12] S. DAWSON, A. BAKHARIA, E. HEATHCOTE, SNAPP: Realising the affordances of real-time SNA within networked learning environments. Presented at the *7th International Conference on Networked Learning*, (2010) Aalborg, Denmark.

[13] S. DAWSON, L. MACFADYEN, L. LOCKYER, D. MAZZOCHI-JONES, Using social network metrics to assess the effectiveness of broad based admission practices. *Australasian Journal of Educational Technology*, 27 (2011), 16–27.

[14] B. DRAGULESCU, *Semantic Web Technologies in educational context*, Phd Thesis, Politehnica University Timişoara, Timisoara, Romania, 2012.

[15] J. ELLSON, E. R. GANSNER, E. KOUTSOFIOS, S. C. NORTH, G. WOODHULL, Graphviz and dynagraph—static and dynamic graph drawing tools, *Graph Drawing Software* (Springer), (2004) pp. 127–148, Berlin, Germany.

[16] R. FERGUSON, S. B. SHUM, Social learning analytics: five approaches, Presented at the *2nd International Conference on Learning Analytics and Knowledge*, (2012) Vancouver, Canada.

[17] G. FLOURIS, I. FUNDULAKI, M. MICHOU, G. ANTONIOU, Controlling access to RDF graphs, Presented at *the Future Internet-FIS 2010*, (2010) Berlin, Germany.

[18] L. C. FREEMAN, Centrality in social networks conceptual clarification, *Social Networks*, 1 (1979), 215–239.

[19] D. GAŠEVIĆ, A. ZOUAQ, R. JANZEN, "Choose Your Classmates, Your GPA Is at Stake!": The Association of Cross-Class Social Ties and Academic Performance, *American Behavioral Scientist*, 57 (2013), 1460–1479.

[20] C. HAYTHORNTHWAITE, M. M. KAZMER, J. ROBINS, S. SHOEMAKER, Community development among distance learners: Temporal and technological dimensions, *Journal of Computer-Mediated Communication*, 6 (2000).

[21] K. M. HOVER, M. MUHLHAUSER, Integrating Distributed Discussions in Web 2.0 Applications and Their Integration with Lecture Recordings, Presented at the *IEEE International Symposium on Multimedia*, (2013) Anaheim, CA, USA.

[22] M. DE LAAT, V. LALLY, L. LIPPONEN, R.-J. SIMONS, Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis, *International Journal of Computer-Supported Collaborative Learning*, 2 (2007), 87–103.

[23] O. T. LENNING, L. H. EBBERS, The Powerful Potential of Learning Communities: Improving Education for the Future, *ASHE-ERIC Higher Education Report*, Vol. 26, No. 6, ERIC Clearinghouse on Higher Education, 1999.

[24] L. LOCKYER, E. HEATHCOTE, S. DAWSON, Informing pedagogical action: Aligning learning analytics with learning design, *American Behavioral Scientist*, 57 (2013), 1439–1459.

[25] A. MCAULEY, B. STEWART, G. SIEMENS, D. CORMIER, *The MOOC model for digital practice*, 2010.

[26] S. NAIDU, E–Learning A Guidebook of Principles, *Procedures and Practices*, 2nd Revised Edition, Commonwealth Educational Media Center for Asia, New Delhi, 2006.

[27] S. RETALIS, A. PAPASALOUROS, Y. PSAROMILIGKOS, S. SISCOS, T. KARGIDIS, Towards Networked Learning Analytics–A concept and a tool, Presented at *the Fifth International Conference on Networked Learning*, (2006) Lancaster, UK.

[28] N. SCLATER, Large-scale open source e-learning systems at the Open University UK, Educause Centre for Applied Research, *Research Bulleting 2008*, 2008.

[29] S. B. SHUM, R. FERGUSON, Social Learning Analytics, *Educational Technology & Society*, 15 (2012), 3–26.

[30] G. SIEMENS, Learning analytics: envisioning a research discipline and a domain of practice, Presented at *the 2nd International Conference on Learning Analytics and Knowledge*, (2012) Vancouver, BC, Canada.

[31] E. SIMPERL, Reusing ontologies on the Semantic Web: A feasibility study, *Data & Knowledge Engineering*, 68 (2009), 905–925.

[32] R. T. SPARROWE, R. C. LIDEN, S. J. WAYNE, M. L. KRAIMER, Social networks and the performance of individuals and groups, *Academy of management journal*, 44 (2001), 316–325.

[33] N. YUSOF, A. B. F. MANSUR, Ontology Development of e-Learning Moodle for Social Learning Network Analysis, *Analysis*, 7 (2013), 2050–2055.

[34] N. YUSOF, A. B. MANSUR, M. S. OTHMAN, Ontology of Moodle e-learning system for social network analysis, Presented at *the IEEE Open Systems (ICOS)*, (2011) Langkawi, Malaysia.

Contact addresses:
Bogdan Drăgulescu
Faculty of Electronics and Telecommunications
Politehnica University Timişoara
Blvd. Vasile Parvan, no. 2
300223 Timişoara
România
e-mail: `bogdan.dragulescu@upt.ro`

Marian Bucos
Faculty of Electronics and Telecommunications
Politehnica University Timişoara
Blvd. Vasile Parvan, no. 2
300223 Timişoara
România
e-mail: `marian.bucos@upt.ro`

Radu Vasiu
Faculty of Electronics and Telecommunications
Politehnica University Timişoara
Blvd. Vasile Parvan, no. 2
300223 Timişoara
România
e-mail: `radu.vasiu@upt.ro`

BOGDAN DRAGULESCU is a postdoc researcher at Department of Comunication, Politehnica University of Timişoara, Romania. He received a Ph.D. in Electronics Engineering and Telecommunications in 2012 from Politehnica University Timisoara. His research interests include semantic web, learning analytics, e-Learning platforms and database management.

MARIAN BUCOS is a Lecturer at the Politehnica University of Timisoara, Communications Department, where he teaches subjects related to database systems, programming, and web-based technologies. Marian Bucos received the M.Sc. and Ph.D. degrees in Electronic Engineering from the Politehnica University of Timisoara, Romania, in 2005 and 2008, respectively. His research interests include modeling and analysis of virtual organizations, data analysis, testing and development of e-Learning platforms. He was involved in 15 research or international cooperation projects.

RADU VASIU received his M.Sc. and Ph.D. degrees in Electronics and Telecommunications Engineering from the Politehnica University of Timisoara, Romania in 1982 and 1997, respectively. He is currently Professor at the Faculty of Electronics and Telecommunications Engineering of the Politehnica University of Timisoara. His latest research interests are in e-Learning, multimedia, web technologies, open data and smart cities. He is President of the Research Committee of the Politehnica University Senate and Director of the Multimedia Research Centre. He has published 12 books and over 100 papers presented at different international conferences. He was involved in 28 research or international cooperation projects. Currently, prof. Vasiu acts as President of the International Association for e-Science (IAFeS) that promotes at international level the use of ICT in science and technology.