A Systematic Mapping Study on the Usage of Software Tools for Graphs within the EDM Community

Vladimir Ivančević*
University of Novi Sad, Faculty of Technical Sciences
Trg Dositeja Obradovića 6
21000 Novi Sad, Serbia
dragoman@uns.ac.rs

Ivan Luković
University of Novi Sad, Faculty of Technical Sciences
Trg Dositeja Obradovića 6
21000 Novi Sad, Serbia
ivan@uns.ac.rs

ABSTRACT

The field of educational data mining (EDM) has been slowly expanding to embrace various graph-based approaches to interpretation and analysis of educational data. However, there is a great wealth of software tools for graph creation, visualization, and analysis, both general-purpose and domain-specific, which may discourage EDM practitioners from finding a tool suitable for their graph-related problem. For this reason, we conducted a systematic mapping study on the usage of software tools for graphs in the EDM domain. By analysing papers from the proceedings of previous EDM conferences we tried to understand how and to what end graph tools were used, as well as whether researchers faced any particular challenges in those cases. In this paper, we compile studies that relied on graph tools and provide answers to the posed questions.

Keywords

Systematic Mapping Study, Graphs, Software Tools, Educational Data Mining.

1. INTRODUCTION

The field of educational data mining (EDM) has significantly expanded over the past two decades. It has attracted numerous researchers with various backgrounds around the common goal of understanding educational data through intelligent analysis and using the extracted knowledge to improve and facilitate learning, as well as educational process. In 2010, Romero and Ventura published a comprehensive overview of the field with 306 references [26]. In this review, the authors identified 11 categories of educational tasks, two of which dealt with graph structures (for brevity these will be referred to as graphs): social network analysis (SNA) and developing concept maps. However, the authors noted that these two categories featured a lower number of papers (15 or less references collected). Somewhat different categories of work were presented in another review of EDM [2] but they did not include any explicit references to graphs.

However, since that time, the interest in approaches and technologies utilizing graphs has increased within EDM. In addition to the results of a literature search on the topic, this could be also evidenced by the appearance of the Workshop on Graph-Based Educational Data Mining (G-EDM)[1] in 2014. As a result, software tools that help researchers or any other user group to utilize graphs or graph-based structures (for brevity these will be referred to as graph tools) are becoming a valuable resource for both the G-EDM and the broader EDM community. As graphs are only slowly gaining wider recognition in EDM, there could still be a lot of questions about which graph tools exist or what educational tasks might be supported by these tools.

In an attempt to help EDM researchers discover more useful information about potentially suitable graph tools, we reviewed the papers presented at the past EDM conferences, selected those that mentioned any usage of graph tools, and extracted from them information about which graph tools the authors employed, what features of these tools were used, to what end the research in question was conducted, and if there were any particular challenges while using these tools.

The present study may be classified as a secondary study since we base our approach on collecting other research works and assembling relevant information from them. Secondary studies might be more typical of medical and social sciences but there are proposed methodologies concerning secondary studies in software engineering as well [13]. Two kinds of secondary studies might be particularly important in this context: systematic review studies and systematic mapping studies [20]. In both cases, there is a clear methodology that is set to reduce bias when selecting other research works, which gives these secondary studies the quality of being systematic. Some of the differences pointed out by Petersen et al. [20] are that systematic reviews tend to focus on the quality of reviewed studies with the aim of identifying best practices, while systematic maps focus more on classification and thematic analysis but with less detailed evaluation of collected studies. Moreover, the same authors consider that the two study types form a continuum, which might complicate some attempts at categorization.

We categorize the present study as a systematic mapping study. This classification is justified by the fact that:

1. we employed a concrete methodology,
2. we did not evaluate the quality of collected papers or the presented results, but
3. we focused on identifying the employed graph tools and the manner in which these tools were used, with the aim

*Corresponding Author

[1] http://ceur-ws.org/Vol -1183/

of providing an overview of the current practice of using graph tools within the EDM community

However, we did not restrict our investigation to analysing exclusively titles, abstracts, or keywords, but went through the complete texts to find the necessary information. This aspect might better suit systematic reviews, but it does not change the principal goal or character of our study.

The exact details of the employed methodology, including the research questions, sources of studies, and study selection criteria, are given in Section 2. Section 3 contains the answers to the research question, most importantly the list of identified graph tools and the trends in their usage in EDM. Section 4 covers the potential limitations of the present study.

2. METHODOLOGY

We mainly followed the guidelines given in [20] but also relied on the example of a mapping study presented in [21]. Given the specificity of our study and the posed research questions, there were some necessary deviations from the standard suggested procedure. The overall process of selecting papers and extracting information, together with the resolution methods for non-standard cases, is presented and discussed in the following subsections.

2.1 Overview

The first step was defining research questions to be answered by the present study. The choice of research questions influenced the subsequent steps: conducting the search for papers, screening the papers, devising the classification scheme, extracting data, and creating a map.

2.2 Research Questions

We defined four principal research questions (RQ1-RQ4) concerning the use of graphs and graph tools in studies by EDM researchers:

- RQ1: Which graph tools were directly employed by researchers in their studies?
- RQ2: Which features of the employed graph tools were used by researchers?
- RQ3: What was the overall purpose of the research that involved or relied on graph tools?
- RQ4: What features did researchers consider to be missing or inadequate in the employed graph tools?

2.3 Search for Papers

We searched through all the papers that were published in the proceedings of the EDM conference series till this date, i.e., papers from the first EDM conference in 2008 to the latest, seventh, EDM conference in 2014. The latest EDM conference was special because it also included four workshops (G-EDM being one of them) for the first time. The papers from these workshops were also considered in our search. This amounted to eight relevant conference proceedings that represented the complete source of research works for our study:

1. Proceedings of the 1st International Conference on Educational Data Mining 2008 (Montreal, Canada)
2. Proceedings of the 2nd International Conference on Educational Data Mining 2009 (Cordoba, Spain)
3. Proceedings of the 3rd International Conference on Educational Data Mining 2010 (Pittsburgh, Pennsylvania, USA)
4. Proceedings of the 4th International Conference on Educational Data Mining 2011 (Eindhoven, Netherlands)
5. Proceedings of the 5th International Conference on Educational Data Mining 2012 (Chania, Greece)
6. Proceedings of the 6th International Conference on Educational Data Mining 2013 (Memphis, Tennessee, USA)
7. Proceedings of the 7th International Conference on Educational Data Mining 2014 (London, UK)
8. Extended Proceedings of the 7th International Conference on Educational Data Mining 2014 (London, UK), which included only the workshop papers

All the proceedings are freely offered as PDF files by the International Society of Educational Data Mining[2] and may be accessed through a dedicated web page.[3]

The papers from these proceeding represented our Level 0 (L0) papers, i.e., the starting set of 494 papers. This set included different categories of papers: full (regular) papers, short papers, different subcategories of posters, as well as works from the young researcher track (YRT) or demos/interactive events. The starting set did not include abstracts of invited talks (keynotes), prefaces of proceedings, or workshop summaries.

These papers were then searched and evaluated against our keyword criterion (KC), which led to a set of Level 1 (L1) papers. Our keyword string is of the form KC1 AND KC2 where KC1 and KC2 are defined in the following manner:

- KC1: graph OR subgraph OR clique
- KC2: tool OR application OR software OR framework OR suite OR package OR toolkit OR environment OR editor

The first part of the criterion (KC1) was defined to restrict the choice to papers that dealt with graphs, while the second part (KC2) served to narrow down the initial set of papers to those mentioning some kind of a tool or program in general.

When evaluating KC on each L0 paper, we did a case-insensitive search for whole words only, whether in their singular form (as written in KC1 and KC2) or their plural form (except for the case of “software”). This search also included hyphenated forms that featured one of the keywords from KC, e.g., “sub-graph” was considered to match the “graph” keyword.

As each proceedings file is a PDF document, we implemented a search in the Java programming language using the Apache PDFBox[4] library for PDF manipulation in Java. However, when extracting content from some papers, i.e., page ranges of a proceedings file, we could not retrieve text in English that could be easily searched. This was most probably caused by the fact that

[2] http ://www.educationaldatamining.org/

[3] http://www.educationaldatamining.org/proceedings

[4] https://pdfbox.apache.org/

authors used different tools to produce camera ready versions in PDF, which were later integrated into a single PDF file.

In these instances, usually one of the two main problems occurred: no valid text could be extracted or valid text was extracted but without spacing. In the case of invalid text, we had to perform optical character recognition (OCR) on the problematic page ranges. We used the OCR feature of PDF-XChange Viewer,[5] which was sufficient as confirmed by our manual inspection of the problematic page ranges (six problematic papers in total). In the case of missing spacing, we had to fine-tune the extraction process using the capabilities of the PDFBox library.

This PDF library proved adequate for our task because we had to search only through PDF files and could customize the text extraction process to solve the spacing problem. However, in the case of a more varied data source, a more advanced toolkit for content indexing and analysis would be needed.

2.4 Screening of Papers

EDM researchers used many of our keywords with several different meanings, e.g., a graph could denote a structure consisting of nodes and edges, which was the meaning that we looked for, or some form of a plot. In order to determine the final set of papers we performed a two-phase selection on L1 papers:

1. We examined the portions of L1 papers that contained some KC1 keyword and eliminated papers that did not significantly deal with graphs (as structures) – this led to a set of Level 2 (L2) papers.
2. We read each L2 paper and eliminated those that did not mention some use of graphs tools – this led to the final set of Level 3 (L3) papers.

In the first phase of selection, we examined the sentences that contain KC1 keywords. If this proved insufficient to determine the nature or scope of use of the mentioned graphs, we read the whole paragraph, and sometimes even the paragraph before and the paragraph after. In these cases, we also checked the referenced figures, tables, or titles of the cited papers. If there were still any doubts, we consulted the paper’s title and abstract, as well as glanced over the figures looking for graph examples. If the authors did not use graphs in their presented study or just made a short comment about graphs giving an analogy or mentioning graphs in the context of related or future work, we did not select the paper for the next phase.

In the second phase of selection, we kept only those papers that mention explicit use of a graph tool by the authors. In the cases when the actual use of a mentioned graph tool was not clear, the paper was selected if some of its figures contain a screenshot featuring the tool or a graph visualized using that tool.

The term tool was considered rather broadly in the present study. We did not restrict the search only to well-rounded software applications, but also included libraries for various computer languages, and even computer languages or file formats that were used by researchers to manipulate graphs. By making this decision, we aimed to provide a greater breadth of information to researchers interested in applying graphs within their studies.

[5] http://www.tracker-software.com/product/pdf-xchange-viewer

2.5 Classification Scheme

The mode of tool usage was categorized in the following manner:

1. CREATION (C) – the tool was developed by the paper authors and introduced in the paper;
2. MODIFICATION (M) – the tool being modified, either through source code or by adding extensions/plugins; and.
3. UTILIZATION (U) – the tool being utilized without modification.

We also checked the distribution of the collected studies by the continent and the country corresponding to the authors’ affiliation. In cases when there were authors from different countries, we indicated the country of the majority of authors, or, if there was no majority then the country corresponding to the affiliation of the first author.

2.6 Data Extraction and Map Creation

Relevant data from L3 papers was extracted into a table that for each paper included the following information: author list, title, proceedings where it was published, page range within the proceedings, answers to the research question and classifications according to the scheme presented in the previous subsection.

3. RESULTS AND DISCUSSION

An overview of the paper selection process is given in Table 1. In each step, the number of relevant papers is significantly reduced. As expected, the required effort in paper analysis was inversely proportional to the number of selected papers. In the L1 step, the usage of the keyword criterion relatively quickly eliminated many papers. However, in subsequent steps, the selected papers had to be read, either partially (in the L2 step) or fully (in the L3 step). The set of L3 papers represents a selection of EDM studies that were used to identify the usage patterns concerning graph tools. The list of the selected papers is publicly available.[6]

Table 1. The number of selected papers at each step

Step	Number of papers
L0 – papers from EDM proceedings	494
L1 – papers containing keywords	146
L2 – papers mentioning graphs	82
L3 – papers mentioning graph tools	27

Most studies (15) are from North America: USA (14) and Canada (1). Europe is represented by 8 studies from 6 countries: Czech Republic (2), Spain (2), Germany (1), Ireland (1), Russia (1), and UK (1). The remaining two continents represented are Asia (Japan only) and Australia, each providing 2 studies. This somewhat resembles the EDM community present at the EDM conferences and differs little from the structure of the EDM community as reported in 2009 [2].

3.1 Overview of Graph Tools

In Table 2, we list 28 graph tools mentioned in the 27 selected papers.

[6] http://www.acs.uns.ac.rs/en/user/31

Table 2. Overview of graph tools from the selected papers

No	Tool	Usage	Features	Purpose	Issues
1	<Untitled framework>	C[1]	argument database – retrieval and mining	retrieve, analyse, and reuse arguments	WIP
2	<Untitled tool>	C[25]	vis. and mine visit trails from WBESs	discover student trails in WBESs	/
3	AGG Engine	C[14], U[15]	augmented graph grammar engine with recursive graph matching	analyse student-produced argument diagrams	inefficiency in some cases
4	CASSI	C[19]	collect bullying data via web-form and use them to form a social graph	support classroom management	/
5	CLOVER framework	U[25]	generate graph vis.	(used in vis. in No. 2)	/
6	Cmate	U[16]	provide a list of concepts and linking words to build a concept map	tabletop concept mapping	/
7	D3.js	U[17]	program interactive graph vis.	facilitate graph interpretation in EDA	/
8	DOT	U[28]	describe graphs	(used in export in No. 14)	/
9	EDM Vis	C[9], M[10]	interactively vis. ITS log data	understand student problem solving in ITSs	WIP
10	eJUNG lib.	U[11]	layout graphs	(used in vis. in No. 14)	/
11	FuzzyMiner (ProM)	U[16]	generate fuzzy models (of student collaboration processes)	discover and analyse student strategies in tabletop collaboration	/
12	Gephi	U[7]	vis. graphs	identify similarities between LE course content (used together with No. 22)	/
13	graphML	U[30]	describe graphs (of student resolution proofs)	analyse student solutions of resolution proofs	/
14	InVis	C[11], M[12, 28]	interactively vis. and edit ITS log data	understand student interaction in ITSs	WIP
15	LeMo	C[18]	interactively vis. learning object networks	understand how students perform and succeed with resources in LMSs and LPs	/
16	Meerkat-ED toolbox	C[22]	vis., monitor, and evaluate participation of students in discussion forums	analyse student interaction and messages in discussion forums	/
17	meud	U[24]	create diagrams (concept lattices)	analyse choices of study programmes	/
18	Ora	U[6]	calculate SNA metrics	study SNA metrics to improve student performance classifiers	/
19	pajek	U[3],[32]	vis. networks and calculate network measures	use student social data to predict drop-out and failure; understand growth of communities on SNSs	/
20	R	U[8]	use scripts to vis. ELE interaction data	explore ELE interaction data and improve ELEs	WIP
21	R – igraph package	U[5],[32]	create, refine, vis., and analyse networks	compare student problem solving-approaches in ITSs; understand growth of communities on SNSs	/
22	RapidMiner	M[7]	create an operator for graph generation	identify similarities between LE course content (used together with No. 12)	/
23	RSP	C[4]	discover issues in the ITS process	support teachers through AT adaptation	/
24	SEMILAR toolkit	C[27]	semantic similarity methods for text	assess student natural language input in ITSs	/
25	SketchMiner	C[29]	generate graphs for student symbolic drawings; compare and cluster drawings	assess student symbolic drawings in ITSs	/
26	STG	C[4]	interactively vis. student interaction in ITSs	understand student problem solving in ITSs	/
27	TRADEM	C[23]	perform analysis on content corpus and generate a concept map in ITSs	support development of instructional content in ITSs	/
28	Visone	U[31]	vis. and analyse SNs, clique analysis	analyse user relationships in WBATs	/

The rows (graph tools) are ordered alphabetically by the tool name (the "Tool" column), which represents the answer to RQ1. In general, we discovered a diverse list of infrequently used graph tools. The usage of the graph tools, which represents the answer to RQ2, is covered by the columns "Usage" and "Features". In "Usage", we listed the mode of usage (see Section 2.5) and the references to the papers mentioning the graph tool. In "Features", we listed tool functionalities and capabilities that were created or employed by the researchers. The most often used feature was to visualize (vis.) graphs. The purpose of the selected studies, which represents the answer to RQ3, is given in the "Purpose" column. Researchers often analysed data from various interrelated systems: intelligent tutoring systems (ITSs) and adaptive tutorials (ATs), learning environments (LEs) including exploratory learning environments (ELEs), learning management systems (LMSs), learning portals (LPs), social network services (SNSs), web-based authoring tools (WBATs), and web-based educational systems (WBESs). Some frequent tasks were analysis of social networks (SNs) and exploratory data analysis (EDA).

The issues that the researchers faced when using the tools, which represents the answer to RQ4, are listed in the "Issues" column. In the majority of the selected papers, the researchers did not discuss problems related to tool usage. The main exceptions are studies in which researcher presented their own tools and discussed missing or incomplete features that should be fully implemented in future – this was labelled as work in progress (WIP).

4. POTENTIAL LIMITATIONS

The findings might not be representative of the whole EDM community but only of the practitioners who presented their work at one of the EDM conferences. An important issue in the analysis was the lack of information about the used tools. There were various instances when researchers obviously used a graph tool, or at least it could be expected that they relied on such tools, but failed to report the information.

Moreover, we used a somewhat "relaxed" definition of a graph tool. This allowed for the inclusion of both general-purpose tools for graph manipulation and domain-specific tools that were developed for educational domain but also utilize a graph-based structure. The primary motive behind this choice was to provide a list of graph tools potentially usable in a wider range of studies, as well as a list of tools that illustrates how graphs were implemented or used in a more specific problem. The former tool category generally includes tools associated with the "U" usage (tools utilized without modification), while the latter tool category mostly covers tools associated with the "C" usage (new tools introduced by their authors).

On the other hand, we excluded graph-based tools that could be labelled as data mining tools or causal modelling tools. For instance, some popular predictive and/or explanatory models (decision trees, random forests, and Bayesian networks) are graph-based, while causal modelling usually assumes creation or discovery of causal graphs. As these tools are more often featured in EDM studies, we assumed that EDM researchers are more familiar with their usage, so the focus of the present study is on other less frequently used graph tools.

5. CONCLUSION

We hope that the collected information about the usage of graph tools within the EDM community may prove valuable for researchers considering the use of graphs to solve educational problems. For future work, we plan to include other publication series, even those that are not solely devoted to the EDM research. The results of such an attempt could demonstrate whether EDM practitioners from other regions of the world are more represented in the graph-based research than indicated by the results of the present study.

6. ACKNOWLEDGMENTS

The research presented in this paper was supported by the Ministry of Education, Science, and Technological Development of the Republic of Serbia under Grant III-44010: "Intelligent Systems for Software Product Development and Business Support based on Models".

7. REFERENCES

[1] Abbas, S. and Sawamura, H. 2008. Argument mining using highly structured argument repertoire. In *Proceedings of the First International Conference on EDM* (Montreal, Canada, June 20 - 21, 2008). EDM'08. 202-209.

[2] Baker, R.S.J.d. and Yacef, K. 2009. The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining,* 1 (1), 3-17.

[3] Bayer, J., Bydzovska, H., Geryk, J., Obsivac, T. and Popelinsky, L. 2012. Predicting drop-out from social behaviour of students. In *Proceedings of the Fifth International Conference on EDM* (Chania, Greece, June 19 - 21, 2012). EDM'12. 103-109.

[4] Ben-Naim, D., Bain, M. and Marcus, N. 2009. A user-driven and data-driven approach for supporting teachers in reflection and adaptation of adaptive tutorials. In *Proceedings of the Second International Conference on EDM* (Cordoba, Spain, July 1 - 3, 2009). EDM'09. 21-30.

[5] Eagle, M. and Barnes, T. 2014. Exploring differences in problem solving with data-driven approach maps. In *Proceedings of the Seventh International Conference on EDM* (London, UK, July 4 - 7, 2014). EDM'14. 76-83.

[6] Garcia-Saiz, D., Palazuelos, C. and Zorrilla, M. 2014. The predictive power of SNA metrics in education. In *Proceedings of the Seventh International Conference on EDM* (London, UK, July 4 - 7, 2014). EDM'14. 419-420.

[7] Goslin, K. and Hofmann, M. 2013. Identifying and visualizing the similarities between course content at a learning object, module and program level. In *Proceedings of the Sixth International Conference on EDM* (Memphis, TN, USA, July 6 - 9, 2013). EDM'13. 320-321.

[8] Gutierrez-Santos, S., Mavrikis, M., Poulovassilis, A. and Zhu, Z. 2014. Indicator visualisation for adaptive exploratory learning environments. In *Proceedings of the Seventh International Conference on EDM* (London, UK, July 4 - 7, 2014). EDM'14. 377-378.

[9] Johnson, M.W. and Barnes, T. 2010. EDM visualization tool: Watching students learn. In *Proceedings of the Third International Conference on EDM* (Pittsburgh, PA, USA, June 11 - 13, 2010). EDM'10. 297-298.

[10] Johnson, M.W., Eagle, M.J., Joseph, L. and Barnes, T. 2011. The EDM Vis tool. In *Proceedings of the Fourth*

International Conference on EDM (Eindhoven, The Netherlands, July 6 - 8, 2011). EDM'11. 349-350.

[11] Johnson, M.W, Eagle, M. and Barnes, T. 2013. InVis: An interactive visualization tool for exploring interaction networks. In *Proceedings of the Sixth International Conference on EDM* (Memphis, TN, USA, July 6 - 9, 2013). EDM'13. 82-89.

[12] Johnson, M.W., Eagle, M., Stamper, J. and Barnes, T. 2013. An algorithm for reducing the complexity of interaction networks. In *Proceedings of the Sixth International Conference on EDM* (Memphis, TN, USA, July 6 - 9, 2013). EDM'13. 248-251.

[13] Kitchenham, B. and Charters, S. 2007. *Guidelines for Performing Systematic Literature Reviews in Software Engineering.* Technical Report EBSE-2007-01, School of Computer Science and Mathematics, Keele University.

[14] Lynch, C.F. 2014. AGG: Augmented graph grammars for complex heterogeneous data. In *Extended Proceedings of the Seventh International Conference on EDM* (London, UK, July 4 - 7, 2014). EDM'14. 37-42.

[15] Lynch, C.F. and Ashley, K.D. 2014. Empirically valid rules for ill-defined domains. In *Proceedings of the Seventh International Conference on EDM* (London, UK, July 4 - 7, 2014). EDM'14. 237-240.

[16] Martinez-Maldonado, R., Yacef, K. and Kay, J. 2013. Data mining in the classroom: discovering groups' strategies at a multi-tabletop environment. In *Proceedings of the Sixth International Conference on EDM* (Memphis, TN, USA, July 6 - 9, 2013). EDM'13. 121-128.

[17] McTavish, T.S. 2014. Facilitating graph interpretation via interactive hierarchical edges. In *Extended Proceedings of the Seventh International Conference on EDM* (London, UK, July 4 - 7, 2014). EDM'14. 59-61.

[18] Merceron, A., Schwarzrock, S., Elkina, M., Pursian, A, Beuster, L., Fortenbacher, A., Kappe, L. and Wenzlaff B. 2013. Visual exploration of interactions and performance with LeMo. In *Proceedings of the Sixth International Conference on EDM* (Memphis, TN, USA, July 6 - 9, 2013). EDM'13. 396-397.

[19] Olson, R., Daily, Z., Malayny, J. and Szkutak, R. 2013. Project CASSI: A social-graph based tool for classroom behavior analysis and optimization. In *Proceedings of the Sixth International Conference on EDM* (Memphis, TN, USA, July 6 - 9, 2013). EDM'13. 398-399.

[20] Petersen, K., Feldt, R., Mujtaba, S. and Mattsson, M. 2008. Systematic mapping studies in software engineering. In *Proceedings of the Twelfth International Conference on Evaluation and Assessment in Software Engineering* (Bari, Italy, June 26 - 27, 2008).

[21] Portillo-Rodriguez, J., Vizcaino, A., Piattini, M. and Beecham, S. 2012. Tools used in Global Software Engineering: A systematic mapping review. *Information and Software Technology.* 54 (Mar. 2012), 663-685.

[22] Rabbany Khorasgani R., Takaffoli, M. and Zaiane, O.R. 2011. Analyzing participation of students in online courses using social network analysis techniques. In *Proceedings of the Fourth International Conference on EDM* (Eindhoven, The Netherlands, July 6 - 8, 2011). EDM'11. 21-30.

[23] Ray, F, Brawner, K. and Robson, R. 2014. Using data mining to automate ADDIE. In *Proceedings of the Seventh International Conference on EDM* (London, UK, July 4 - 7, 2014). EDM'14. 429-430.

[24] Romashkin, N., Ignatov, D. and Kolotova, E. 2011. How university entrants are choosing their department? Mining of university admission process with FCA taxonomies. In *Proceedings of the Fourth International Conference on EDM* (Eindhoven, The Netherlands, July 6 - 8, 2011). EDM'11. 230-233.

[25] Romero, C., Gutierrez, S., Freire, M. and Ventura, S. 2008. Mining and visualizing visited trails in web-based educational systems. In *Proceedings of the First International Conference on EDM* (Montreal, Canada, June 20 - 21, 2008). EDM'08. 182-186.

[26] Romero, C. and Ventura, S. 2010. Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews,* 40 (6), 601-618.

[27] Rus, V., Banjade, R., Lintean, M., Niraula, N. and Stefanescu, D. 2013. SEMILAR: A semantic similarity toolkit for assessing students' natural language inputs. In *Proceedings of the Sixth International Conference on EDM* (Memphis, TN, USA, July 6 - 9, 2013). EDM'13. 402-403.

[28] Sheshadri, V., Lynch, C. and Barnes, T. 2014. InVis: An EDM tool for graphical rendering and analysis of student interaction data. In *Extended Proceedings of the Seventh International Conference on EDM* (London, UK, July 4 - 7, 2014). EDM'14. 65-69.

[29] Smith, A., Wiebe, E., Mott, B. and Lester, J. 2014. SKETCHMINER: Mining learner-generated science drawings with topological abstraction. In *Proceedings of the Seventh International Conference on EDM* (London, UK, July 4 - 7, 2014). EDM'14. 288-291.

[30] Vaculik, K., Nezvalova, L. and Popelinsky, L. 2014. Graph mining and outlier detection meet logic proof tutoring. In *Extended Proceedings of the Seventh International Conference on EDM* (London, UK, July 4 - 7, 2014). EDM'14. 43-50.

[31] Xu, B. and Recker, M.M. 2010. Peer production of online learning resources: A social network analysis. In *Proceedings of the Third International Conference on EDM* (Pittsburgh, PA, USA, June 11 - 13, 2010). EDM'10. 315-316.

[32] Yamakawa, O., Tagawa, T., Inoue, H., Yastake, K. and Sumiya T. 2011. Combining study of complex network and text mining analysis to understand growth mechanism of communities on SNS. In *Proceedings of the Fourth International Conference on EDM* (Eindhoven, The Netherlands, July 6 - 8, 2011). EDM'11. 335-336.