

A Survey of Applications that use Graph-based Educational Data Mining

João Oliveira¹, Rodrigo Alexandrino¹, Ana Paula Ambrosio¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74001-970 – Goiania– GO – Brazil

{joaosantos, rodrigoalexandrino, apaula}@inf.ufg.br

Abstract. *Graph-based Educational Data Mining can be used to analyze educational scenarios that are best represented by complex networks. To manipulate these networks, several algorithms have been proposed and are implemented in tools that facilitate the development of practical applications. This article presents a survey of applications in the educational domain implemented using graph-based data mining, and published from 2000 to 2019. 30 papers were selected. Information extracted from these papers include the research questions they proposed to answer, the adopted graph representation, and the data mining algorithms and tools used. Non-statistical methods were used to evaluate and interpret the findings. The results highlight the domains and types of problems where graph-based education data mining is being used and the graphs/networks used to represent these contexts, as well as the dominant data mining approaches being adopted to extract information from these structures.*

Resumo. *A Mineração de Dados Educacionais baseada em grafos pode ser usada para analisar cenários educacionais que podem ser melhor representados por redes complexas. Para manipular essas redes, vários algoritmos foram propostos e implementados em ferramentas que facilitam o desenvolvimento de aplicações práticas. Este artigo apresenta um levantamento de aplicações no domínio educacional que implementaram mineração de dados baseados em grafos, e foram publicados entre 2000 e 2019. Ao total 30 artigos foram selecionados. As informações extraídas desses trabalhos incluem as questões de pesquisa que eles propuseram responder, a representação gráfica adotada e os algoritmos e ferramentas de mineração de dados usados. Métodos não estatísticos foram utilizados para avaliar e interpretar os achados. Os resultados destacam os domínios e tipos de problemas em que a mineração de dados baseada em grafos está sendo usada e os gráficos / redes usados para representar esses contextos, bem como as abordagens de mineração de dados dominantes adotadas para extrair informações dessas estruturas.*

1. Introduction

The application of data mining techniques on educational data has developed into a new knowledge domain known as educational data mining [Baker and Yacef 2009]. Traditional techniques work mainly with data structured in record-based formats [Fayyad et al. 1996b]. However, some applications are better represented using complex networks, for example to represent social networks and the relations between

users [Figueiredo 2011]. In order to obtain information from complex networks, graph mining techniques have been used as efficient approaches to knowledge extraction [Chakrabarti and Faloutsos 2006].

With time, the application of complex networks and graph mining techniques has become more popular [London et al. 2015]. Often, these complex systems can be represented by graphs or networks, where the vertices or nodes represent individuals or entities, while the edges or links represent the interaction between pairs of these individuals. The complex network approach is not only useful to simplify and visualize this huge amount of interconnected data, but is also effective in identifying key elements of the system and finding its most important interactions, as well as discovering groupings of network characteristics [Fayyad et al. 1996a]. Several papers describe techniques and theories of network generation and information extraction [Kavitha et al. 2011] [London et al. 2015] [Rehman et al. 2012] [Lima Filho et al.] [London and Németh 2014].

Graph-based Educational Data Mining (G-EDM) has been the new term used by researchers who extract information from compelling networks of data derived from educational bases [Lynch et al. 2015]. This graphical data includes complex logs of user-system interaction or between users, graphs generated by the student, and conceptual hierarchies that have large amounts of graphical data [F Lynch 2017]. There is a wealth of pedagogical information under these graphical data sets. As a result, graphical data mining techniques, such as grammatical induction of graphs, path analysis, and the identification of necessary prerequisites, have become increasingly important.

To date, three international workshops on Graphics-based Educational Data Mining have been successfully proposed within the more general conference on Educational Data Mining (EDM). The first was held in London, co-located with EDM 2014 [Loureiro and Goussevskaia 2015]. It had 12 publications, of which 6 were full-papers, the remaining were short papers. The second was co-located with EDM 2015 in Spain [Lynch et al. 2015]. 10 authors submitted their work for publication, including 4 full papers and 6 short papers. The third was co-located with EDM 2017 in China [F Lynch 2017].

Although there are more papers related to the subject, many do not use the expression "Graph-based Educational Data Mining (G-EDM)" or similar, making them hard to find. This survey aims to identify these papers and group them under one reference that can be used by other researchers in this domain.

In this study, we survey several papers that address practical applications of information extraction in complex networks in very different contexts. It is intended to highlight which are the most appropriate graph mining tools to analyze them and what kind of additional information can be extracted by their use. Also, what are the analysis trends and what researchers are searching, and what kinds of approaches are more frequent. It aims to show how the Graph Mining domain has expanded and what challenges it still has to overcome.

2. Research methodology

Following the original guidelines proposed by [Kitchenham et al. 2009] the main objective of this study is to elaborate a survey of articles that contain practical research in-

volving G-EDM with a special focus on those articles that do not mention the term "G-EDM" but are linked to the domain. A secondary objective is to elucidate expanding concepts and important techniques that benefit the field of education in order to further explore the field of study.

2.1. Research questions

The following research questions are addressed, and are distinguished into primary (generalized: set to fulfill the goals of the review) and secondary (sub-objectives/specific: refine the primary - explanatory).

- Q1: What is the main focus of studies in Graph-based Educational Data Mining?
Q1.1 What are the results and contributions of these studies?
- Q2: Which are the most frequent techniques used to extract information in graphs?

To carry out the literature review, we defined a review protocol, composed of four distinct stages: I) the literature - collect data, II) review and evaluate research results - selection of primary studies, III) analyze, code and synthesize the results, and IV) report the review.

2.2. Search string

For the search process, both automatic and manual search were used on national and international databases. Automatic search was done using string search and keywords. The sites consulted in the research were: *ACM Digital Library*, *IEEE Xplore*, *Google Scholar*, *Elsevier(Scopus and ScienceDirect)* and *SciELO*. The search terms included: *Graph Mining*, *Educational Graph Mining*, *Graph-based Educational Data Mining*, *Social Network Analysis*, *Network Analysis*, *Complex networks analysis*, *Educational Social network analysis*.

Each database has its own search mechanism, so different strings were used to account for these specificities. In the *ACM Digital Library* database the following *string* was used:

("graph mining") AND ("educational") OR ("scholar") OR ("academic") OR ("student") OR ("education")

It returned **24 articles** of which only **2**, published from 2000 to 2019, were selected by the inclusion and exclusion criteria (see Section 2.3).

The *IEEE Xplore* database was queried using the *string*:

(((((graph mining) AND educat) OR scholar) OR student*) OR academic)*

It returned 12,971 papers. To refine the search, only journal were included, and the total was reduced to **32** results from which **5** papers, published between 2000 - 2019, were selected following the inclusion and exclusion criteria.

Thus the automatic search yielded 7 articles. To complement the research, a manual search for articles was undertaken using related terms (Networks, Educational Graph, and SNA educational), on Google Scholar and SciELO, as well as in the Proceedings of G-EDM. This manual search resulted in 23 selected articles. Figure 1 shows the distribution by year of publishing.

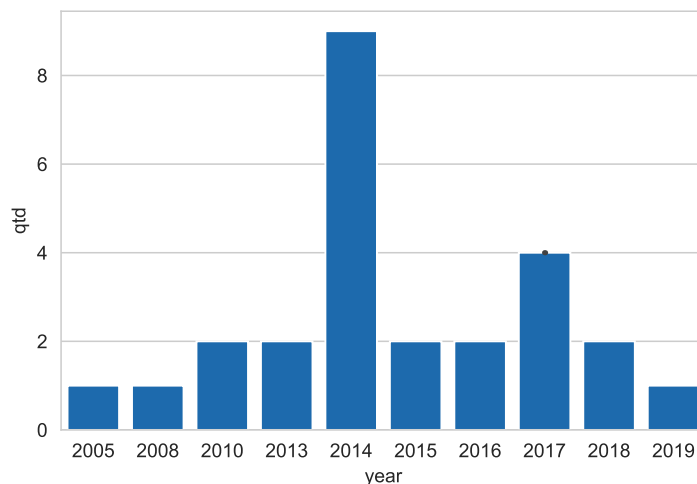


Figura 1. Seleção dos artigos subdivido por ano.

2.3. Inclusion and Exclusion Criteria

Based on the aim of this survey, to identify applications that use G-EDM, inclusion/exclusion criteria were defined.

Include:

- Articles published in Journals or Proceedings.
- Articles that include information extraction from complex networks.
- Full papers and short papers will be accepted.

Exclude:

- Articles that present some analysis with graphs but do not use an educational database.
- Theoretical articles without practical implementation will be excluded.
- Book chapters.

The G-EDM is an emerging domain and educational studies using graph representation are new and many are published as short papers. A total of **30 articles** were selected after systematic analysis and application of the inclusion and exclusion criteria.

3. Results and discussion

Table 1 contains the selected papers. They are classified according to the type of analysis used in the research as defined in the Proceedings of G-EDM [F Lynch 2017]. Articles included in *Graphical behavior models* contain studies that observe the different behaviors of the networks in several temporal stages. *Graph-based log analysis* studies data from interaction in learning environments and virtual interactions. Five papers in this category analyzed LMS logs describing interactions in the environment. *Graphical solution representations* are used to solve problems through network representation. *Novel graph analysis* present new solutions (algorithms or methods). *Tools and technologies for automatic concept hierarchy extraction* focus on graphical solutions of hierarchy (priority

dependence). *Tools and Technologies for Graph grammar (pattern) recognition* are studies that present grammar solutions or novel representation models. *Relevant analytical tools and standard problems* discuss new analytic tools. *Social network analysis* present solutions based on the concept of Social Networks Analysis (SNA).

Tabela 1. Seleção dos trabalhos pela análise pretendida, em alguns casos um artigo pode aparecer em mais de uma análise.

Point of View	Reference
Graphical behavior models	[Okamoto et al. 2014], [Brown et al. 2015], [Kumar 2014], [Jiang et al. 2014]
Graph-based log analysis	[Gitinabard et al. 2017], [Dawson 2008], [Gardner and Brooks 2018], [Grawemeyer et al. 2017], [Kovanovic et al. 2014], [Mostafavi and Barnes 2014], [Yang et al. 2013]
Graphical solution representations	[Yeung et al. 2005], [Li et al. 2015], [Tang et al. 2008], [Gašević et al. 2013], [Eckles and Stradley 2012], [Fire et al. 2012], [Dekel and Gal 2014]
Novel graph analysis techniques	[Buluz and Yilmaz 2017], [Patel et al. 2017], [Costa et al. 2019], [London and Németh 2014], [Zheng et al. 2017]
Tools and technologies for automatic concept hierarchy extraction	[Chen et al. 2018], [Cateté et al. 2014], [Yang et al. 2010]
Tools and technologies for graph grammar (pattern) recognition	[Lynch 2014]
Relevant analytical tools and standard problems	[Sheshadri et al. 2014]
Social network analysis	[Wejnert 2010], [Kovanovic et al. 2014], [Garcia-Saiz et al. 2014]

The *Directed Acyclic Graph* is the most widely used representation (34%), followed by *Directed Acyclic Multigraph* with 23% (see Table 2). In general Directed Acyclic Graph are graphs are often used in analysis of user interactions in virtual environments.

Aggarwal and Wang [Aggarwal and Wang 2010] define three main approaches in graph mining: classification, clustering, and discovery of frequent patterns (or substructures). Table 3 organizes the selected papers according to the adopted approach. Some studies use more than one approach. Classification is the more frequent approach. After analysis, it was identified that most of the studies are related to the behavior and identification of student profiles.

3.1. Discussion of Research Questions

Q1: What is the main focus of research in Graph-based Educational Data Mining? In 21 papers, approximately 70% of the selected papers, research focuses on student analysis, their behavior in virtual environments and student interactions [Okamoto et al. 2014] [Gitinabard et al. 2017] [Dawson 2008] [Li et al. 2015] [Gašević et al. 2013] [Gardner and Brooks 2018] [Brown et al. 2015] [Eckles and Stradley 2012]

Tabela 2. Seleção dos artigos subdividido pelo tipo de grafo.

Types of Graph	References
Directed Acyclic Graph	[Buluz and Yilmaz 2017], [Chen et al. 2018], [Kumar 2014], [Cateté et al. 2014], [Costa et al. 2019], [Wejnert 2010],[London and Németh 2014], [Zheng et al. 2017],[Yang et al. 2013], [Gitinabard et al. 2017]
Directed Cyclic Graph	[Yang et al. 2010]
Directed Acyclic Multigraph	[Dawson 2008], [Yeung et al. 2005], [Li et al. 2015], [Brown et al. 2015], [Patel et al. 2017], [Kovanovic et al. 2014], [Ramos et al. 2016]
Simple Acyclic Graph	[Tang et al. 2008], [Gašević et al. 2013], [Grawemeyer et al. 2017],[Lynch 2014], [Jiang et al. 2014]
Tree	[Gardner and Brooks 2018], [Dekel and Gal 2014], [Mostafavi and Barnes 2014], [Sheshadri et al. 2014]
Undirected Acyclic Multigraph	[Eckles and Stradley 2012], [Fire et al. 2012]
Undirected Hipergraph	[Okamoto et al. 2014]

[Buluz and Yilmaz 2017] [Grawemeyer et al. 2017] [Patel et al. 2017]
 [Fire et al. 2012] [Kovanovic et al. 2014] [Cateté et al. 2014][Jiang et al. 2014]
 [Mostafavi and Barnes 2014] [Sheshadri et al. 2014] [Costa et al. 2019]
 [Yang et al. 2013] [Zheng et al. 2017] [London and Németh 2014][Dekel and Gal 2014].
 One focused on grammar models for learning [Lynch 2014] [Kumar 2014] and the others
 on tools [Tang et al. 2008][Ramos et al. 2016] [Chen et al. 2018], process learning
 and Social Network analysis [Fire et al. 2012] [Yang et al. 2010][Yeung et al. 2005]
 [Gardner and Brooks 2018] [Eckles and Stradley 2012] [Wejnert 2010].

Q1.1 What are the results and contributions of these studies? Several articles concentrate their analysis on the user and aim to promote improvements in the learning process, understand student, teacher and researcher behaviors, build patterns of student profiles and behaviors in social networks, and understand the students' learning process in virtual environments. Some of the selected studies deserve special attention. For example, the KnowEdu system [Chen et al. 2018], which automatically constructs a knowledge graph for education to facilitate understanding of how students build knowledge in virtual environments. The same approach is adopted by [Ramos et al. 2016] [Dekel and Gal 2014] [Yang et al. 2010]. In studies by [Okamoto et al. 2014] [Dawson 2008] [Gašević et al. 2013] [Gardner and Brooks 2018] [Buluz and Yilmaz 2017] [Fire et al. 2012] [Kovanovic et al. 2014] [Yang et al. 2013] the contribution involves the techniques of extracting information from graphs. They include i) learning ii) analysis of students' behavior iii) correlation and association of patterns in students vi) others, through measures of complex networks such as Degree, Betweenness and Centrality degree. In [Zheng et al. 2017] techniques and algorithms of machine learning were applied to identify student behavior in open courses open (MOOCs) as a conclusion it was possible to suggest better techniques of content delivery and distribution of study groups. A similar approach by [Kumar 2014] [Buluz and Yilmaz 2017] [Gardner and Brooks 2018]. The analysis of social networks

Tabela 3. Seleção de artigos pelo tipo de abordagem em mineração de grafos.

Graph mining approach	References
Classification	[Tang et al. 2008], [Gašević et al. 2013], [Fire et al. 2012], [Dekel and Gal 2014], [Kovanovic et al. 2014], [Kumar 2014], [Lynch 2014], [Cateté et al. 2014], [Jiang et al. 2014], [Mostafavi and Barnes 2014], [Sheshadri et al. 2014], [Costa et al. 2019], [Wejnert 2010], [London and Németh 2014], [Zheng et al. 2017], [Yang et al. 2013], [Yang et al. 2010], [Ramos et al. 2016], [Gitinabard et al. 2017], [Dawson 2008], [Brown et al. 2015], [Grawemeyer et al. 2017], [Chen et al. 2018]
Clustering	[Yeung et al. 2005], [Yang et al. 2013]
Discovery of frequent patterns	[Okamoto et al. 2014], [Li et al. 2015], [Brown et al. 2015], [Eckles and Stradley 2012], [Buluz and Yilmaz 2017], [Patel et al. 2017], [Gardner and Brooks 2018]

is another aspect well studied by the authors [Wejnert 2010] [Yeung et al. 2005] [Tang et al. 2008] where it is possible to extract behaviors and characteristics of distinct or common profiles to improve academic performance.

Q2: Which are the most frequent tools used to extract information in graphs?

Most tools focus on analyzing and visualizing the data. NetVis (2D) [Dawson 2008] presents an interface, evaluation metrics and network information extraction algorithms. GEPHI [Gašević et al. 2013] [Costa et al. 2019] is one of the most complete, offering data import, algorithms and data visualization. Neo4J [Grawemeyer et al. 2017] is a graph-based database where it is possible, through its query language Cypher, to construct and visualize graphs from the data stored in the database. LPGraph [Ramos et al. 2016] is a plugin for R Studio. While it is possible to make analysis in graphs, there are few visualization resources and analysis algorithms. Netbeans [Sheshadri et al. 2014] through Invis is a Java tool for viewing and extracting information in graphs. Python NetworkX [Grawemeyer et al. 2017] [Fire et al. 2012] [Lynch et al. 2015] is a Python script for graph analysis and visualization, and offers one of the largest varieties of algorithms for graph analysis. The Rgraph plugin for R [Fire et al. 2012], [Patel et al. 2017] presents some algorithms and metrics for visualization and extraction of information in graphs.

4. Conclusion

This study focused on Graph-based Educational Data Mining applications, but because the term did not exist until 2014, research related to educational data mining using some approach/technique with graphs were included in this survey. Although three workshops have occurred within the more general Educational Data Mining conference, G-EDM is still a little explored domain, that tends to grow with time.

Representation via Graphs has been shown to be an efficient technique of visualization and representation, mainly when it involves databases that clearly present relationships between entities.

The ever growing interest in Education and in understanding the teaching/learning processes are strong promoters of G-EDM. The relevance of social media and online education, where interaction between actors are naturally represented by graph data, lead to research that aim to extract pedagogical information that will contribute to improving academic success. An increase is, therefore, expected for research in this domain.

Referências

- Aggarwal, C. C. and Wang, H. (2010). Graph data management and mining: A survey of algorithms and applications. In *Managing and mining graph data*, pages 13–68. Springer.
- Baker, R. S. and Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM— Journal of Educational Data Mining*, 1(1):3–17.
- Brown, R., Lynch, C., Wang, Y., Eagle, M., Albert, J., Barnes, T., Baker, R. S., Bergner, Y., and McNamara, D. S. (2015). Communities of performance & communities of preference. In *EDM (Workshops)*.
- Buluz, B. and Yilmaz, B. (2017). Graph mining approach for modeling academic success. In *Signal Processing and Communications Applications Conference (SIU), 2017 25th*, pages 1–4. IEEE.
- Cateté, V., Hicks, D., Lynch, C., and Barnes, T. (2014). Snag'em: Graph data mining for a social networking game. In *Workshop on Graph-Based Educational Data Mining*, volume 6, page 10.
- Chakrabarti, D. and Faloutsos, C. (2006). Graph mining: Laws, generators, and algorithms. *ACM computing surveys (CSUR)*, 38(1):2.
- Chen, P., Lu, Y., Zheng, V. W., Chen, X., and Yang, B. (2018). Knowedu: A system to construct knowledge graph for education. *IEEE Access*.
- Costa, J. d. J., Bernardini, F., Artigas, D., and Viterbo, J. (2019). Mining direct acyclic graphs to find frequent substructures—an experimental analysis on educational data. *Information Sciences*, 482:266–278.
- Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Journal of educational technology & society*, 11(3).
- Dekel, R. and Gal, K. (2014). On-line plan recognition in exploratory learning environments. In *EDM (Workshops)*.
- Eckles, J. E. and Stradley, E. G. (2012). A social network analysis of student retention using archival data. *Social Psychology of Education*, 15(2):165–180.
- F Lynch, Collin & Barnes, T. & X. L. & G. N. (2017). Graph-based educational data mining (g-edm 2017) proceedings. In *Proceedings of the 10th International Conference on Educational Data Mining*, pages 1–30.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996a). From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (1996b). Advances in knowledge discovery and data mining.

- Figueiredo, D. R. (2011). Introdução a redes complexas. *Atualizações em Informática*, pages 303–358.
- Fire, M., Katz, G., Elovici, Y., Shapira, B., and Rokach, L. (2012). Predicting student exam's scores by analyzing social network data. In *International Conference on Active Media Technology*, pages 584–595. Springer.
- Garcia-Saiz, D., Palazuelos, C., and Zorrilla, M. (2014). Data mining and social network analysis in the educational field: An application for non-expert users. In *Educational Data Mining*, pages 411–439. Springer.
- Gardner, J. and Brooks, C. (2018). Coenrollment networks and their relationship to grades in undergraduate education. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, pages 295–304. ACM.
- Gašević, D., Zouaq, A., and Janzen, R. (2013). “choose your classmates, your gpa is at stake!” the association of cross-class social ties and academic performance. *American Behavioral Scientist*, 57(10):1460–1479.
- Gitinabard, N., Xue, L., Lynch, C. F., Heckman, S., and Barnes, T. (2017). A social network analysis on blended courses. *arXiv preprint arXiv:1709.10215*.
- Grawemeyer, B., Wollenschlaeger, A., Gutierrez, S., Holmes, W., Mavrikis, M., and Poullovassilis, A. (2017). Using graph-based modelling to explore changes in students' affective states during exploratory learning tasks.
- Jiang, S., Fitzhugh, S. M., and Warschauer, M. (2014). Social positioning and performance in moocs. In *Workshop on Graph-Based Educational Data Mining*, volume 14.
- Kavitha, D., Rao, B. M., and Babu, V. K. (2011). A survey on assorted approaches to graph data mining. *International Journal of Computer Applications*, 14(1):43–46.
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., and Linkman, S. (2009). Systematic literature reviews in software engineering—a systematic literature review. *Information and software technology*, 51(1):7–15.
- Kovanovic, V., Joksimovic, S., Gasevic, D., and Hatala, M. (2014). What is the source of social capital? In *Workshop on Graph-Based Educational Data Mining*, page 1. Citeseer.
- Kumar, R. (2014). Cross-domain performance of automatic tutor modeling algorithms. In *EDM (Workshops)*.
- Li, F., Long, X., Du, S., Zhang, J., Liu, Z., Li, M., Li, F., Gui, Z., and Yu, H. (2015). Analyzing campus mobility patterns of college students by using gps trajectory data and graph-based approach. In *Geoinformatics, 2015 23rd International Conference on*, pages 1–5. IEEE.
- Lima Filho, S. P., Cavalcanti, M. C., and Justel, C. M. Modelagem em grafos a partir de bancos de dados relacionais.
- London, A. and Németh, T. (2014). Student evaluation by graph based data mining of administrative systems of education. In *Proceedings of the 15th International Conference on Computer Systems and Technologies*, pages 363–369. ACM.

- London, A., Pelyhe, Á., Holló, C., and Németh, T. (2015). Applying graph-based data mining concepts to the educational sphere. In *Proceedings of the 16th International Conference on Computer Systems and Technologies*, pages 358–365. ACM.
- Loureiro, A. A. F. and Goussevskaia, O. (2015). Grafos. *Disponível em: [http://homepages.dcc.ufmg.br/loureiro/md/md Grafos. pdf](http://homepages.dcc.ufmg.br/loureiro/md/md%20Grafos.pdf)*. Acessado em, 20.
- Lynch, C. F. (2014). Agg: Augmented graph grammars for complex heterogeneous data. In *EDM (Workshops)*.
- Lynch, C. F., Barnes, T., Albert, J., and Eagle, M. (2015). Graph-based educational data mining (g-edm 2015). *CEUR-WS*.
- Mostafavi, B. and Barnes, T. (2014). Evaluation of logic proof problem difficulty through student performance data. In *EDM (Workshops)*. Citeseer.
- Okamoto, K., Asanuma, H., and Kawamoto, K. (2014). A graph based data mining method for collaborative learning space in learning commons. In *World Automation Congress (WAC), 2014*, pages 415–420. IEEE.
- Patel, N., Sellman, C., and Lomas, D. (2017). Mining frequent learning pathways from a large educational dataset. *arXiv preprint arXiv:1705.11125*.
- Ramos, D. B. et al. (2016). Uma ferramenta baseada em grafo para identificação e visualização de trilhas de aprendizagem.
- Rehman, S. U., Khan, A. U., and Fong, S. (2012). Graph mining: A survey of graph mining techniques. In *Digital Information Management (ICDIM), 2012 Seventh International Conference on*, pages 88–92. IEEE.
- Sheshadri, V., Lynch, C., and Barnes, T. (2014). Invis: An edm tool for graphical rendering and analysis of student interaction data. In *EDM (Workshops)*.
- Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., and Su, Z. (2008). Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 990–998. ACM.
- Wejnert, C. (2010). Social network analysis with respondent-driven sampling data: A study of racial integration on campus. *Social Networks*, 32(2):112–124.
- Yang, C., Park, H., and Heo, J. (2010). A network analysis of interdisciplinary research relationships: The korean government’s r&d grant program. *Scientometrics*, 83(1):77–92.
- Yang, D., Sinha, T., Adamson, D., and Rosé, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-driven education workshop*, volume 11, page 14.
- Yeung, Y.-Y., Liu, T. C.-Y., and Ng, P.-H. (2005). A social network analysis of research collaboration in physics education. *American Journal of Physics*, 73(2):145–150.
- Zheng, Y., Liu, R., and Hou, J. (2017). The construction of high educational knowledge graph based on mooc. In *Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2017 IEEE 2nd Information*, pages 260–263. IEEE.