

# Graph-based Educational Data Mining (G-EDM 2015)

Collin F. Lynch  
Department of Computer  
Science  
North Carolina State  
University  
Raleigh, North Carolina  
cflynch@ncsu.edu

Dr. Jennifer Albert  
Department of Computer Science  
North Carolina State University  
Raleigh, North Carolina  
jlsharp@ncsu.edu

Dr. Tiffany Barnes  
Department of Computer  
Science  
North Carolina State  
University  
Raleigh, North Carolina  
tmbarnes@ncsu.edu

Michael Eagle  
Department of Computer  
Science  
North Carolina State  
University  
Raleigh, North Carolina  
mjeagle@ncsu.edu

## 1. INTRODUCTION

Fundamentally, a graph is a simple concept. At a basic level a graph is a set of relationships  $\{e(n_0, n_2), e(n_0, n_j), \dots, e(n_{j-1}, n_j)\}$  between elements. This simple concept, however, has afforded the development of a complex theory of graphs [1] and rich algorithms for combinatorics [7] and clustering [4]. This has, in turn, made graphs a fundamental part of educational data mining.

Many types of data can be naturally represented as graphs such as social network data, user-system interaction logs, argument diagrams, logical proofs, and forum discussions. Such data has grown exponentially in volume as courses have moved online and educational technology has been incorporated into the traditional classroom. Analyzing it can help to answer a range of important questions such as:

- What path(s) do high-performing students take through online educational materials?
- What social networks can foster or inhibit learning?
- Do users of online learning tools behave as the system designers expect?
- What diagnostic substructures are commonly found in student-produced diagrams?
- Can we use prior student data to identify students' solution plan, if any?
- Can we use prior student data to provide meaningful hints in complex domains?
- Can we identify students who are particularly helpful based upon their social interactions?

Thus, graphs are simple in concept, general in structure, and have wide applications for Educational Data Mining (EDM). Despite the importance of graphs to data mining and data analysis there exists no strong community of researchers focused on Graph-Based Educational Data Mining. Such a community is important to foster useful interactions, share tools and techniques, and to explore common problems.

## 2. GEDM 2014

This is the second workshop on Graph-Based Educational Data Mining. The first was held in conjunction with EDM 2014 in London [17]. The focus of that workshop was on seeding an initial community of researchers, and on identifying shared problems, and avenues for research. The papers presented covered a range of topics including unique visualizations [13], social capital in educational networks [8], graph mining [19, 11], and tutor construction [9].

The group discussion sections at that workshop focused on the distinct uses of graph data. Some of the work presented focused on student-produced graphs as solution representations (e.g. [14, 3]) while others focused more on the use of graphs for large-scale analysis to support instructors or administrators (e.g. [18, 13]). These differing uses motivate different analytical techniques and, as participants noted, change our underlying assumptions about the graph structures in important ways.

## 3. GEDM 2015

Our goal in this second workshop was to build upon this nascent community structure and to explore the following questions:

1. What common goals exist for graph analysis in EDM?
2. What shared resources such as tools and repositories are required to support the community?
3. How do the structures of the graphs and the analytical methods change with the applications?

The papers that we include here fall into four broad categories: interaction, induction, assessment, and MOOCs.

Work by Poulouvassilis et al. [15] and Lynch et al. [12] focuses on analyzing user-system interactions in state based learning environments. Poulouvassilis et al. focuses on the analyses of individual users' solution paths and presents a novel mechanism to query solution paths and identify general solution strategies. Lynch et al. by contrast, examined user-system interactions from existing model-based tutors to examine the impact of specific design decisions on student performance.

Price & Barnes [16] and Hicks et al. [6] focus on applying these same analyses in the open-ended domain of programming. Unlike more discrete tutoring domains where users enter single equations or select actions, programming tutors allow users to make drastic changes to their code on each step. This can pose challenges for data-driven methods as the student states are frequently unique and admit no easy single-step advice. Price and Barnes present a novel method for addressing the data sparsity problem by focusing on minimal-distance changes between users [16] while in related work Hicks et al. focuses on the use of path weighting to select actionable advice in a complex state space [6].

The goal in much of this work is to identify rules that can be used to characterize good and poor interactions or good and poor graphs. Xue et al. sought address this challenge in part via the automatic induction of graph rules for student-produced diagrams [22]. In their ongoing work they are applying evolutionary computation to the induction of Augmented Graph Grammars, a graph-based formalism for rules about graphs.

The work described by Leo-John et al. [10], Guerra [5] and Weber & Vas [21], takes a different tack and focuses not on graphs representing solutions or interactions but on relationships. Leo-John et al. present a novel approach for identifying closely-related word problems via semantic networks. This work is designed to support content developers and educators in examining a set of questions and in giving appropriate assignments. Guerra takes a similar approach to the assessment of users' conceptual changes when learning programming. He argues that the conceptual relationship graph affords a better mechanism for automatic assessment than individual component models. This approach is also taken up by Weber and Vas who present a toolkit for graph-based self-assessment that is designed to bring these conceptual structures under students' direct control.

And finally, Vigentini & Clayphan [20], and Brown et al. [2] focus on the unique problems posed by MOOCs. Vigentini and Clayphan present work on the use of graph-based metrics to assess students' on-line behaviors. Brown et al., by contrast, focus not on local behaviors but on social networks with the goal of identifying stable sub-communities of users and of assessing the impact of social relationships on users' class performance.

## 4. REFERENCES

- [1] B. Bollobás. *Modern Graph Theory*. Springer Science+Business Media Inc. New York, New York, U.S.A., 1998.
- [2] R. Brown, C. F. Lynch, Y. Wang, M. Eagle, J. Albert, T. Barnes, R. Baker, Y. Bergner, and D. McNamara. Communities of performance & communities of preference. In C. F. Lynch, T. Barnes, J. Albert, and M. Eagle, editors, *Proceedings of the Second International Workshop on Graph-Based Educational Data Mining (GEDM 2015)*. CEUR-WS, June 2015. (in press).
- [3] R. Dekel and K. Gal. On-line plan recognition in exploratory learning environments. In S. G. Santos and O. C. Santos, editors, *Proceedings of the Workshops held at Educational Data Mining 2014, co-located with 7th International Conference on Educational Data Mining (EDM 2014)*, London, United Kingdom, July 4-7, 2014., volume 1183 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.
- [4] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. of the National Academy of Sciences*, 99(12):7821–7826, June 2002.
- [5] J. Guerra. Graph analysis of student model networks. In C. F. Lynch, T. Barnes, J. Albert, and M. Eagle, editors, *Proceedings of the Second International Workshop on Graph-Based Educational Data Mining (GEDM 2015)*. CEUR-WS, June 2015. (in press).
- [6] A. Hicks, V. Catete, R. Zhi, Y. Dong, and T. Barnes. Bots: Selecting next-steps from player traces in a puzzle game. In C. F. Lynch, T. Barnes, J. Albert, and M. Eagle, editors, *Proceedings of the Second International Workshop on Graph-Based Educational Data Mining (GEDM 2015)*. CEUR-WS, June 2015. (in press).
- [7] D. E. Knuth. *The Art of Computer Programming: Combinatorial Algorithms, Part 1*, volume 4A. Addison-Wesley, 1<sup>st</sup> edition, 2011.
- [8] V. Kovanovic, S. Joksimovic, D. Gasevic, and M. Hatala. What is the source of social capital? the association between social network position and social presence in communities of inquiry. In S. G. Santos and O. C. Santos, editors, *Proceedings of the Workshops held at Educational Data Mining 2014, co-located with 7th International Conference on Educational Data Mining (EDM 2014)*, London, United Kingdom, July 4-7, 2014., volume 1183 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.
- [9] R. Kumar. Cross-domain performance of automatic tutor modeling algorithms. In S. G. Santos and O. C. Santos, editors, *Proceedings of the Workshops held at Educational Data Mining 2014, co-located with 7th International Conference on Educational Data Mining (EDM 2014)*, London, United Kingdom, July 4-7, 2014., volume 1183 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.
- [10] R.-J. Leo-John, T. McTavish, and R. Passonneau. Semantic graphs for mathematics word problems based on mathematics terminology. In C. F. Lynch, T. Barnes, J. Albert, and M. Eagle, editors, *Proceedings of the Second International Workshop on Graph-Based Educational Data Mining (GEDM 2015)*. CEUR-WS, June 2015. (in press).
- [11] C. F. Lynch. AGG: augmented graph grammars for complex heterogeneous data. In S. G. Santos and O. C. Santos, editors, *Proceedings of the Workshops held at Educational Data Mining 2014, co-located with 7th International Conference on Educational Data Mining (EDM 2014)*, London, United Kingdom, July 4-7, 2014., volume 1183 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.
- [12] C. F. Lynch, T. W. Price, M. Chi, and T. Barnes. Using the hint factory to analyze model-based tutoring systems. In C. F. Lynch, T. Barnes, J. Albert, and M. Eagle, editors, *Proceedings of the Second International Workshop on Graph-Based Educational Data Mining (GEDM 2015)*. CEUR-WS, June 2015. (in press).
- [13] T. McTavish. Facilitating graph interpretation via interactive hierarchical edges. In S. G. Santos and O. C. Santos, editors, *Proceedings of the Workshops held at Educational Data Mining 2014, co-located with 7th International Conference on Educational Data Mining (EDM 2014)*, London, United Kingdom, July 4-7, 2014., volume 1183 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.
- [14] B. Mostafavi and T. Barnes. Evaluation of logic proof problem difficulty through student performance data. In S. G. Santos

- and O. C. Santos, editors, *Proceedings of the Workshops held at Educational Data Mining 2014, co-located with 7th International Conference on Educational Data Mining (EDM 2014)*, London, United Kingdom, July 4-7, 2014., volume 1183 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.
- [15] A. Poulouvassilis, S. G. Santos, and M. Mavrikis. Graph-based modelling of students' interaction data from exploratory learning environments. In C. F. Lynch, T. Barnes, J. Albert, and M. Eagle, editors, *Proceedings of the Second International Workshop on Graph-Based Educational Data Mining (GEDM 2015)*. CEUR-WS, June 2015. (in press).
- [16] T. Price and T. Barnes. An exploration of data-driven hint generation in an open-ended programming problem. In C. F. Lynch, T. Barnes, J. Albert, and M. Eagle, editors, *Proceedings of the Second International Workshop on Graph-Based Educational Data Mining (GEDM 2015)*. CEUR-WS, June 2015. (in press).
- [17] S. G. Santos and O. C. Santos, editors. *Proceedings of the Workshops held at Educational Data Mining 2014, co-located with 7th International Conference on Educational Data Mining (EDM 2014)*, London, United Kingdom, July 4-7, 2014, volume 1183 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.
- [18] V. Sheshadri, C. Lynch, and T. Barnes. Invis: An EDM tool for graphical rendering and analysis of student interaction data. In S. G. Santos and O. C. Santos, editors, *Proceedings of the Workshops held at Educational Data Mining 2014, co-located with 7th International Conference on Educational Data Mining (EDM 2014)*, London, United Kingdom, July 4-7, 2014., volume 1183 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.
- [19] K. Vaculík, L. Nezvalová, and L. Popelínský. Graph mining and outlier detection meet logic proof tutoring. In S. G. Santos and O. C. Santos, editors, *Proceedings of the Workshops held at Educational Data Mining 2014, co-located with 7th International Conference on Educational Data Mining (EDM 2014)*, London, United Kingdom, July 4-7, 2014., volume 1183 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.
- [20] L. Vigentini and A. Clayphan. Exploring the function of discussion forums in moocs: comparing data mining and graph-based approaches. In C. F. Lynch, T. Barnes, J. Albert, and M. Eagle, editors, *Proceedings of the Second International Workshop on Graph-Based Educational Data Mining (GEDM 2015)*. CEUR-WS, June 2015. (in press).
- [21] C. Weber and R. Vas. Studio: Ontology-based educational self-assessment. In C. F. Lynch, T. Barnes, J. Albert, and M. Eagle, editors, *Proceedings of the Second International Workshop on Graph-Based Educational Data Mining (GEDM 2015)*. CEUR-WS, June 2015. (in press).
- [22] L. Xue, C. F. Lynch, and M. Chi. Graph grammar induction by genetic programming. In C. F. Lynch, T. Barnes, J. Albert, and M. Eagle, editors, *Proceedings of the Second International Workshop on Graph-Based Educational Data Mining (GEDM 2015)*. CEUR-WS, June 2015. (in press).