Graph-based Educational Data Mining (G-EDM 2017)

Collin F. Lynch, Tiffany Barnes, Linting Xue, & Niki Gitinabard
North Carolina State University, Raleigh, North Carolina, USA
cflynch, tmbarnes, lxue3, & ngitina@ncsu.edu

ABSTRACT

With the growing popularity of MOOCs and computer-aided learning systems, as well as the growth of social networks in education, we have begun to collect increasingly large amounts of educational graph data. This graph data includes complex user-system interaction logs, student-produced graphical representations, and conceptual hierarchies that large amounts of graph data have. There is abundant pedagogical information beneath these graph datasets. As a result, graph data mining techniques such as graph grammar induction, path analysis, and prerequisite relationship prediction has become increasingly important. Also, graphical model techniques (e.g. Hidden Markov Models or probabilistic graphical models) has become more and more important to analyze educational data.

While educational graph data and data analysis based on graphical models has grown increasingly common, it's necessary to build a strong community for educational graph researchers. This workshop will provide such a forum for interested researchers to discuss ongoing work, share common graph mining problems, and identify technique challenges. Researchers are encouraged to discuss prior analyses of graph data and educational data analyses based on graphical models. We also welcome discussions of in-progress work from researchers seeking to identify suitable sources of data or appropriate analytical tools.

1. PRIOR WORKSHOPS

So far, we have successfully held two international workshops on Graph-based Educational Data-Mining. The first one was held in London, co-located with EDM 2014. It featured 12 publications of which 6 were full-papers, the remainder short papers. Having roughly 25 full-day attendees and additional drop-ins, it led to a number of individual connections between researchers and the formation of an e-mail list for group discussion. The second one was co-located with EDM 2015 in Spain. 10 authors presented their published work including 4 full papers and 6 short papers there.

2. OVERVIEW AND RELEVANCE

Graph-based data mining and educational data analysis based on graphical models have become emerging disciplines in EDM. Large-scale graph data, such as social network data, complex user-system interaction logs, student-produced graphical representations, and conceptual hierarchies, carries multiple levels of pedagogical information. Exploring such data can help to answer a range of critical questions such as:

- For social network data from MOOCs, online forums, and user-system interaction logs:
  - What social networks can foster or hinder learning?
  - Do users of online learning tools behave as we expect them to?
  - How does the interaction graph evolve over time?
  - What data we can use to define relationship graphs?
  - What path(s) do high-performing students take through online materials?
  - What is the impact of teacher-interaction on students' observed behavior?
  - Can we identify students who are particularly helpful in a course?
- For computer-aided learning (writing, programming, etc.):
  - What substructures are commonly found in student-produced diagrams?
  - Can we use prior student data to identify students' solution plan, if any?
  - Can we automatically induce empirically-valid graph rules from prior student data and use induced graph rules to support automated grading systems?

Graphical model techniques, such as Bayesian Network, Markov Random Field, and Conditional Random Field, have been widely used in EDM for student modeling, decision making, and knowledge tracing. Utilizing these approaches can help to:

- Learn students' behavioral patterns.
- Predict students' behaviors and learning outcomes.

- Induce pedagogical strategies for computer-aided learning systems.
- Identify the difficult level of the knowledge components in the intelligent tutoring systems.

Researches related to these questions can help us to better understand students' learning status, and improve the teaching effectiveness and student learning. Our goal in this workshop is to bring together researchers with special interest in graph-based data analysis to 1) discuss state of the art tools and technologies, 2) identify common problems and challenges, and 3) foster a community of researchers for further collaboration. We will consider the submission of full and short papers as well as posters and demonstrations covering a range of graphics topics that include, but are not limited to:

- Social network data
- Graphical solution representations
- Graphical behavior models
- Graph-based log analysis
- Large network datasets
- Novel graph-based machine learning methods
- Novel graph analysis techniques
- Relevant analytical tools and standard problems
- Issues with graph models
- Tools and technologies for graph grammar (pattern) recognition
- Tools and technologies for automatic concept hierarchy extraction
- Computer-aided learning system development involved with graphical representations
- Use of graphical models in educational data

We particularly welcome submissions of in-progress work both from students and researchers with problems who are seeking appropriate analytical tools, and developers of graph analysis tools who are seeking new challenges.

3. WORKSHOP ORGANIZERS

Dr.Collin F. Lynch is an Assistant Professor in the Department of Computer Science at North Carolina State University. His primary research is focused on graph-based educational data mining, the development of robust intelligent tutoring systems, and adaptive educational systems for ill-defined domains such as scientific writing, law, and engineering. In his more recent work he has also been involved in the development of Intelligent Tutoring Systems for Logic and Probability and social networking analysis for research communities.

Dr.Tiffany Barnes is an Associate Professor of Computer Science at NC State University. She received an NSF-CAREER Award for her novel work in using data to add intelligence to STEM learning environments. That grant supported the development of InVis a novel tool that use graph-based representations of student-tutor interaction data to evaluate the impact of intelligent tutoring systems on student problem-solvers and to automatically extract hints and student advice from log data using graph-analysis. More recently she has received grants for the analysis of large-scale online courses and the development of procedural guidance from intelligent tutoring system data.

Linting Xue is a third year Ph.D. student in the Department of Computer Science at North Carolina State University. She is interested in the graph data mining methods for educational graph data. Her current research is focused on automatically graph grammars induction for student-produced argument diagrams. The induced graph grammars can be used as features for automatic grading and provide the hints for argumentative writing.

Niki Gitinabard is a second year Ph.D. student in the Department of Computer Science at North Carolina State University. She is interested in social network analysis in learning environments. She is currently working on social graph generation and analysis based on students' explicit and implicit interactions.

4. WORKSHOP ORGANIZATION

We will organize this workshop as a full or half-day mini-conference with time set aside for paper presentations, large-group discussion, and individual networking. We will open the workshop with a summary of prior meetings. We will spend the morning on presentations with a short discussion session before lunch. The afternoon session will be divided between presentations and working groups which will focus on identifying shared problems, small-group networking, and planning for follow up work. We will invite submissions of full papers which describe mature work. We will also accept short papers describing in-progress work or student projects, and poster/demo submissions for those presenting available data, tools, and methods. This last category is particularly targeted at researchers who have data or methods available and are seeking to identify potential collaborators.