# Workshop on Graph-Based Educational Data Mining (G-EDM)

Graph data has become increasingly prevalent in data-mining and data analysis generally. Many types of data can be represented naturally as graphs including social network data, log traversal, and online discussions. Moreover recent work on the importance of social relationships, peer tutoring, collaboration, and argumentation has highlighted the importance of relational information in education including:

- Graphical solution representations such as argument diagrams and concept maps;
- Graph-based models of problem-solving strategies;
- User-system interaction data in online courses and open-ended tutors;
- Sub-communities of learners, peer-tutors and project teams within larger courses; and
- Class assignments within a larger knowledge space.

Our goal in this workshop was to highlight the importance of graph data and its relevance to the wider EDM community. We also sought to foster the development of an interested community of inquiry to share common problems, tools, and techniques. We solicited papers from academic and industry professionals focusing on: common problems, analytical tools, and established research. We also particularly welcomed new researchers and students seeking collaboration and guidance on future directions. It is our hope that the papers published here will serve as a foundation for ongoing research in this area and as a basis for future discussions.

The papers included here cover a range of topics. Kovanovic, Joksimovic, Gasevic & Hatala focus on evaluating social networks, and specifically on the development of social capital and high-status individuals in a course context while Catete, Hicks, Barnes, & Lynch describe an online tool designed to promote social network formation in new students. Similar work is also described by Jiang, Fitzhugh & Warschauer who focus on the identification of high-connection users in MOOCs.

Other authors turned to the extraction of plan and hint information from course materials and user logs. Belacel, Durand, & Laplante define a graph-based algorithm for identifying the best path through a set of learning objects. Kumar describes an algorithm for the automatic construction of behavior graphs for example-tracing tutors based upon expert solutions and Dekel & Gal in turn consider plan identification to support automatic guidance. Two further papers by Vaculík, Nezvalová & Popelínský, and by Mostafavi & Barnes, apply graph analysis techniques to the specific domain of logic tutoring and, in particular, on the classification of student solutions and to the evaluation of problem quality.

And finally several authors chose to present general tools for the evaluation of graphical data. Lynch describes Augmented Graph Grammars, a formal rule representation for the analysis of rich graph data such as argument diagrams and interconnected student assignments, and details an implementation of it. Sheshadri, Lynch, & Barnes present InVis a visualization and analysis platform for student interaction data designed to support the types of research described above. And

McTavish describes a general technique to support graph analysis and visualization particularly for student materials through the use of interactive hierarchical edges. We thank the included authors for their contributions to the discussion and look forward to continued research.

**The G-EDM workshop organizers**
*Collin F. Lynch*
*Tiffany M. Barnes*