Towards Data-Informed Group Formation Support Across Learning Spaces

Ishari Amarasinghe, Davinia Hernández-Leo, Anders Jonsson
ICT Department, Universitat Pompeu Fabra, Barcelona, Spain
{ishari.amarasinghe, davinia.hernandez, anders.jonsson}@upf.edu

Abstract: Learning via collaboration has gained much success over past few decades given their learning benefits. Group composition has been seen as a relevant design element that contributes to the potential effectiveness of collaborative learning. To support practitioners in this context this paper addresses the problem of automatic group formation implementing policies related to well-known collaboration techniques and considering personal attributes in across-spaces contexts where multiple activities, places and tools are involved in a learning situation. Analytics of contextual and progress-in-activity information about learners presented as a summary would support practitioners to obtain a comprehensive knowledge about them to subsequently facilitate formation of effective collaborative groups to face forthcoming activities. The paper discusses a work in progress web based architecture of a group formation service to compute groupings which also assists in recommending grouping constraints via learning analytics which will facilitate practitioners in the adaptive set-up of the group formation design element across-spaces.

Keywords: learning analytics, computer supported collaborative learning, collaborative group formation, jigsaw, social learning

Introduction

Over the past few decades research conducted in different disciplines have confirmed active collaborative learning is an effective means of instruction which can be utilized in both traditional and online educational environments that would result in long-term effects in education (Redmond, 2001). Group work conducted under proper conditions provides an opportunity for students to clarify and refine their understanding of concepts through discussions and rehearsals with peers (Konert, Burlak & Steinmetz, 2014; Christodoulopoulos & Papanikolaou, 2007). However, learning via interactions does not occur in every situation (Cruz & Isotani, 2014). Careful consideration over the design of collaboration is as key to achieve desired learning goals.

With the advancements in web technologies and social media students collaborate with each other not only in the physical classroom spaces defined by the formal educational contexts, but also across different digital spaces. In such a context computer supported collaborative learning (CSCL) could effectively mediates interactions among distant learners and co-present learners via computer-based scripts supporting uninterrupted collaboration irrespective of learner's physical location. Or students can engage in flows or sequences of pedagogically-interconnected collaborative learning activities, each proposing a different group formation policy and supported using a different digital collaboration tool (Manathunga & Hernández-Leo, 2016).

However, designing and implementing interconnected flows of activities using different learning spaces are not straightforward. For instance in online learning spaces like MOOCs where thousands of students get registered for a particular course or in large classroom cohorts with, for example, over a hundred (or even less) students it becomes difficult and time-consuming for practitioners to go through each learner's profile or / and actions in previous activities in the flow in order to decide which grouping parameter, or combination of parameters, are pedagogically interesting to be considered in group formation policies and calculate the groupings accordingly (Balmaceda, Schiaffino, & Pace, 2014; Ounnas, Davis, & Millard, 2008). Hence researchers have been investigating several techniques to automate the process of group formation via Computer Supported Group Formation (CSGF) (Balmaceda, Schiaffino, & Pace, 2014; Ounnas, Davis, & Millard, 2008) which provides computational support to complete group formation task successfully. However, existing approaches do not focus on solving across-spaces learning situations where parameters for group formation

policies come from constraints depending on the pedagogical method behind the flow of activities but also from students' characteristics and their monitored behavior and performance during the flow.

Considering these across-spaces learning situations needs, in this paper we describe a work-in-progress web based architecture of a group formation service called "IGroups" which automates learner group formation and employs methods from learning analytics to provide glimpse towards understanding what occurs in different learning spaces to facilitate the identification of relevant parameters for group formation for forthcoming activities in a flow of pedagogically interconnected tasks and tools. This will aid practitioners not only to overcome time consuming group formation tasks but also to design and monitor the space of collaboration.

Related Work

Many studies have pointed out that formation of well-structured collaborative learning groups as the starting point of CSCL (Cruz & Isotani, 2014; Kardan & Sadeghi, 2014). One major approach of forming student groups is based on considering different factors related to student profiles (Sanz-Martinez et al., 2015; Manathunga & Hernández-Leo, 2016). Grouping learners with different learning profiles results in heterogeneous groups while members who are similar to one another can be grouped together forming homogeneous groups (Dillenbourg & Tchounikine, 2007). Further, multiple constraints defined by an educator towards a collaboration task or constraints inherited from a Collaborative Learning Flow Pattern (CLFP) behind a pedagogical method may also become important when formulating student groups (Manathunga & Hernández-Leo, 2016). Groups formed without careful consideration often causes problems such as disproportionate participation of individuals, demotivation and resistance to group work in future activities (Dillenbourg & Tchounikine, 2007; Cruz & Isotani, 2014).

There has been an increasing number of prior works in the field of CSGF. Different algorithmic approaches have been suggested over time to formulate student groups using different approaches (Abnar, Orooji, & Taghiyareh, 2012; Sun & Shen, 2014; Wang, Lin, & Sun, 2007; Kardan & Sadeghi, 2014; Ounnas, Davis, & Millard, 2008). Among some of the efforts towards in which authors describes initial efforts towards web-based group formation systems include DIANA (Wang, Lin, & Sun, 2007), OptAssign (Meyer, 2009) and groupformation.org (Henry, 2013). In DIANA (Wang, Lin, & Sun, 2007) learner group formation was carried out prioritizing student's personal tendencies and attitudes associated when using their own skills to formulate heterogeneous groups. In OptAssign (Meyer, 2009) group formation was modelled as a family of assignment problems. They have reported the evaluation results but have concluded highlighting the requirement of better analytical tools to investigate the quality of the solutions obtained. Moreover, in groupformation.org (Henry, 2013) student information was gathered via a preliminary survey and a preference survey which was then used to create student profiles. Further, homogeneous and heterogeneous student groups based on instructor defined criteria was facilitated. However, authors have not provided experimental results of the suggested approach.

Requirements towards an across-space data informed group formation support

During the literature review it was noticed that aforementioned systems do not appear to be deployed for real classroom usage for practitioners. Existing systems do not take the advantage of connecting heterogeneous data sources which will provide significant insights towards how learning occurs across different spaces. Although in many situations practitioners have access towards an enormous amount of student data, knowledge which could be extracted from this data is left untouched due to barriers in technical expertise. In some situations, learner data spread across heterogeneous sources (e.g., log files, form responses, assessment marks, survey results, lab/library attendance data, demographics etc..) might require a considerable amount of time to process manually.

On the other hand, it was noticed that different authors suggest (Henry, 2013) to carry out preliminary surveys to capture student data with respect to different criteria before forming collaborative learner groups. In our perspective, this will create an additional burden on instructors since they have to design and share additional surveys prior group work. If students' responses are delayed grouping activity will also be delayed and it was noticed that authors have not discussed how to incorporate incomplete survey results and its effect towards grouping criteria. Surprisingly it was noticed that these systems do not take the full advantage of the digital age meaning that they do not incorporate already collected data and automatically tracked data rightly available across different digital spaces rather they wait and restrict the systems to a preliminary survey. In such a context, it is of importance to leverage powerful learning analytics which would be advantageous for practitioners during different phases of collaborative sessions as follows.

Firstly, during the design phase of a collaborative learning activity, learning analytics could provide a broader insight towards learners as a summary. These types of analytics for instance would help practitioners when deciding which pedagogical approaches will best suit for students in a particular learning environment.

Further, clustering algorithms such as K-Means can be used to partition student's data, providing practitioners hints towards deciding extrinsic/soft constraints which best fits for group formation in a particular context. Secondly, during the run time of a collaborative learning task, learning analytics could provide insights towards engagement and behavioral patterns of individual students. Further, it could also help in identifying students who are having less engagement or problems during collaborations. This information will make aware practitioners about students who require personalized support and assistance. Finally, after finishing a collaborative learning task, learning analytics could provide reflections (Verbert et al., 2014) on learning occurred supporting better decision making in future sessions.

IGroups System Architecture

The IGroups system will be implemented adapting to common three-tier web architecture including presentation, logic and data tier. Main objectives of this system development are twofold; firstly, it automates the process of assigning students to collaborative groups based on different policies (heterogeneity, homogeneity, CLFPs), secondly it provides useful and significant insights in determining possible factors that will guide collaboration towards success via learning analytics module.

Formulation of collaborative learner groups was implemented using constraint optimization techniques using a novel binary integer programming approach (Amarasinghe, 2016) adhering to CLFPs (i.e., Jigsaw, Pyramid) which will pre-structure collaboration (Hernández-Leo et al., 2005) based on constraints defined for group formation (Manathunga & Hernández-Leo, 2016). Further, regrouping of students while adapting to changes occur in the learning environments are also facilitated.

Since, the learning analytics module will be implemented to obtain the maximum advantage of using student data which spans across heterogeneous sources it was determined to integrate "IGroups" system to other existing third party software systems via application programming interfaces e.g., REST API. These third-party software systems may include well known and widely used educational platforms such as Moodle LMS, social media platforms or other tools supporting the activities in a learning flow. Student data spread across heterogeneous sources will be processed and presented for practitioners. Computation of designed groupings according to the decided parameters will be also presented for practitioners for their refinement, if required, and accessible via another API for the automatic setup of grouping configurations in tools to be used in forthcoming activities. Investigation of which types of learning analytics as well as visualizations might be of interest in day today teaching practices to support learning design is another research area yet to be explored (Michos & Hernández-Leo, 2016), as a first step we decided to provide learning analytics using easy to interpret visualizations. These visualizations would provide useful hints and guidance towards practitioners on deciding criteria for group formation on demand using readily available information. This information will not be limited only towards basic knowledge which could be extracted via learner's profiles such as demographics but will also include summarized information on their previous performance levels, collaborative behavior during past peer interactions, social communication and interactions across different digital spaces.

Data informed group formation support in across-spaces example

This example demonstrates how practitioners could carry out design and implementation data informed collaborative learning activities via "IGroups" system. Assume a scenario that the instructor wants to carry out a collaborative learning activity in a research methods course at a master's program class. Major objectives of this collaborative task is to familiarize students with the existing research groups in the University with the goal of helping students to identify faculty members who could guide and collaborate during their master thesis. The time duration given to finish the collaborative task was limited to three weeks. Since it is important to consider student's research interests before allocating them to study a particular research group instructors may decide to use learning analytics module in the IGroups system. At this point system, will extract individual student's interests from profile records available in the LMS database and will be presented towards practitioner supporting them to make data driven decisions on how to allocate students to study a particular research group at the University.

Further, instructor may also consider students previous research experiences since mixing of less / no experienced students with experienced students will promote helping among themselves. Research experiences could be extracted via information publicly available in student's LinkedIn profiles. This information will then be presented as statistical summaries towards practitioner which will be useful on deciding feasibility towards formulation of balanced groups based on previous research experiences.

After deciding on grouping criteria instructor will also decide on a CLFP to carry out collaborative tasks. Assuming the instructor would like to formulate collaborative groups based on widely adopted Jigsaw

CLFP given its benefits he/she will utilize the group formation functionality implemented in the IGroups system. Based on Jigsaw CLFP at the expert phase instructor wants to allocate students who are having similar research interests to study the same research group which matches best with their interests. It is also decided to mix student's based on previous research experience levels given its benefits. And at the next stage of Jigsaw
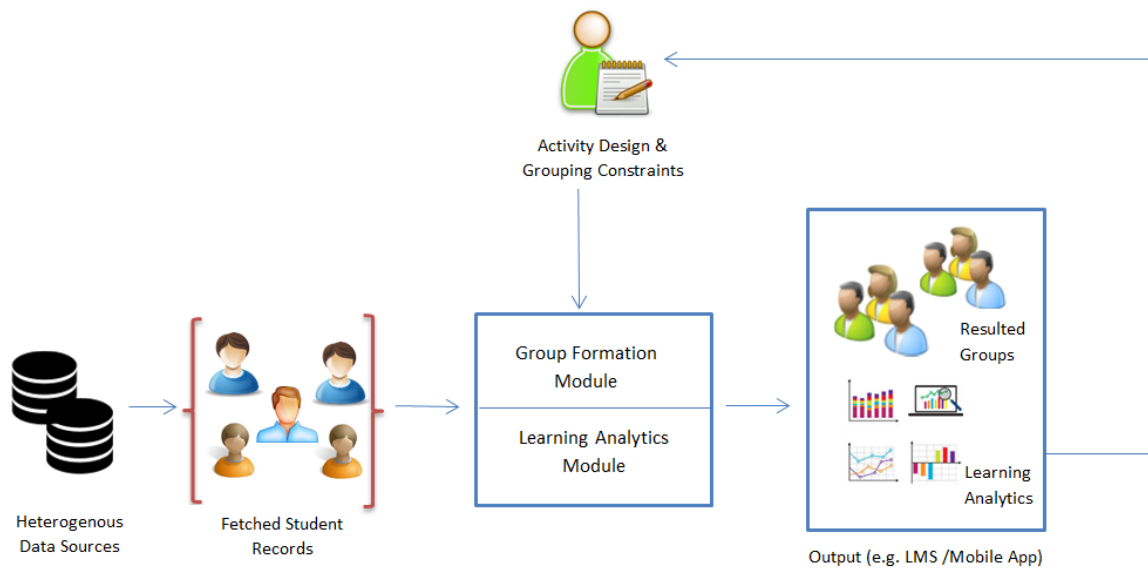

Figure 1. IGroups system high-level design

CLFP it is decided to allocate students who have studied different research groups to the same Jigsaw group, hence sharing knowledge within Jigsaw groups will enhance each other's awareness towards different research groups at the University. After deciding on the aforementioned grouping structures instructor could utilize the functionality implemented in the IGroups system for calculating optimal student groups based on the criteria specified. Group allocations will be then communicated to students via Moodle LMS.

During expert phase of Jigsaw CLFP students who are allocated to study on a particular research group will meet with faculty members to get to know their ongoing research and research focuses. Students are advised to share knowledge gathered via discussions in the group twitter account using a particular hashtag. Students in the same group can comment on interesting research carried out by different faculty members or re tweet peer's posts which they think is important. While students are engaged in the collaborative activity instructors can monitor student's engagement and interactions during the task with the help of learning analytics module in the IGroups system which will provide analytics after analyzing tweets that matched the specified hash tag. For instance, instructors could revisit student's weekly participation in the collaborative task based on analytics generated considering total number of tweets, retweets and comments made. These analytics could also be shared with students providing information on how other groups are engaged in the collaborative task. This type of sharing could increase student motivation and engagement. Further, instructors will also be presented with student clusters based on group performance. Easy to understand visualizations which also facilitates some interactivity which demonstrates how changing of group structures would affect performance levels would provide hints for instructors to decide on grouping criteria (based on performance during expert phase) which needs to be adhered during Jigsaw phase.

Instructor will then input grouping criteria to formulate Jigsaw groups to the IGroups system and Jigsaw group allocations will be communicated to the students. At the end of the Jigsaw activity each student will rate their interests towards working with a particular research group during their master thesis via a Moodle mobile application. This data will then be processed and presented via learning analytics module of IGroups system providing insights on whether collaborative activity has resulted in fruitful outcomes.

Conclusions

This paper describes a work in progress architecture of a web based group formation system which supports educational practitioners when formulating collaborative learner groups while taking into account existing student's data spans across heterogeneous tools, and sources. It is an architecture with open programming interfaces for its integration with data sources (academic systems, educational tools, etc.) and collaboration tooling relevant to support learning activities. Adaptive collaboration is supported via flexible computerized scripts which enables practitioners when handling changes occur in the collaborative space. Further, learning analytics incorporated into group formation service will provide practitioners useful insights during different stages (at the beginning, progress-in activity, post activity) of a collaborative task. Such insights would help practitioners to make data driven decisions towards more potentially effective student's groupings for the setting up of different tools supporting multiple tasks involved in a flow of collaborative learning activities.

References

Abnar, S., Orooji, F., & Taghiyareh, F. (2012). An evolutionary algorithm for forming mixed groups of learners in web based collaborative learning environments. In *Technology Enhanced Education (ICTEE), 2012 IEEE International Conference on* (pp. 1-6). IEEE.

Amarasinghe, I. (2016). Intelligent Group Formation in Compter Supported Collaborative Learning Scripts (Master's Thesis, Universitat Pompeu Fabra, Barcelona, Spain).

Balmaceda, J. M., Schiaffino, S. N., & Pace, J. A. D. (2014). Using constraint satisfaction to aid group formation in CSCL. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial, 17*(53), 35-45.

Christodoulopoulos, C. E., & Papanikolaou, K. A. (2007). A group formation tool in an e-learning context. In *Tools with Artificial Intelligence (ICTAI), 2007 19th IEEE International Conference on* (pp. 117-123). IEEE.

Cruz, W.M., & Isotani, S. (2014) Group Formation Algorithms in Collaborative Learning Contexts: A Systematic Mapping of the Literature. In: Baloian N., Burstein F., Ogata H., Santoro F., & Zurita G. (Eds.), *Collaboration and Technology. CRIWG 2014. Lecture Notes in Computer Science* (pp. 199-214). Springer International Publishing.

Dillenbourg, P., & Tchounikine, P. (2007). Flexibility in macro-scripts for computer-supported collaborative learning. *Journal of computer assisted learning, 23*(1), 1-13.

Henry, T. R. (2013). Creating effective student groups: An introduction to groupformation. org. In *Proceeding of the 44th ACM technical symposium on Computer science education* (pp. 645-650). ACM.

Hernández-Leo, D., Asensio-Pérez, J. I., Dimitriadis, Y., Bote-Lorenzo, M. L., Jorrín-Abellán, I. M., & Villasclaras-Fernández, E. D. (2005). Reusing IMS-LD formalized best practices. *Advanced Technology for Learning*, *2*(3), 223–232.

Kardan, A. A., & Sadeghi, H. (2014). Modeling the learner group formation problem in computer-supported collaborative learning using mathematical programming. In *e-Learning and e-Teaching (ICeLeT), 2014 8th National and 5th International Conference on* (pp. 1-5). IEEE.

Konert, J., Burlak, D., & Steinmetz, R. (2014). The group formation problem: an algorithmic approach to learning group formation. In *European Conference on Technology Enhanced Learning* (pp. 221-234). Springer International Publishing.

Manathunga, K., & Hernández-Leo, D. (2016). A Multiple Constraints Framework for Collaborative Learning Flow Orchestration. In *International Conference on Web-Based Learning* (pp. 225-235). Springer.

Meyer, D. (2009). OptAssign - A web-based tool for assigning students to groups. *Computers & Education*, *53*(4), 1104-1119.

Michos, K., & Hernández-Leo, D. (2016). Towards understanding the potential of teaching analytics within educational communities. In *Proceeding of the 4th International Workshop on Teaching Analytics, IWTA 2016, European Conference on Technology Enhanced Learning, Lyon, France.*

Ounnas, A., Davis, H., & Millard, D. (2008). A framework for semantic group formation. In *Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on* (pp. 34-38). IEEE.

Redmond, M. A. (2001). A computer program to aid assignment of student project groups. In *Proceedings of the thirty-second SIGCSE technical symposium on Computer Science Education* (pp. 134-138). New York, USA.

Sanz-Martinez, L., Ortega-Arranz, A., Dimitriadis, Y., Munoz-Cristobal, J. A., Martinez-Mones, A., Bote-Lorenzo, M. L., & Rubia-Avi, B. (2015). *Identifying factors that affect team formation and management in MOOCS.* Retrieved from https://www.gsic.uva.es/uploaded_files/77623_[ITS2016WS]%20Sanz-

Martinez%20et%20al.pdf

Sun, G., & Shen, J. (2014). Facilitating social collaboration in mobile cloud-based learning: A teamwork as a service (TaaS) approach. *IEEE Transactions on Learning Technologies*, *7*(3), 207–220.

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing, 18*(6), 1499-1514.

Wang, D. Y., Lin, S. S. J., & Sun, C. T. (2007). DIANA: A computer-supported heterogeneous grouping system for teachers to conduct successful small learning groups. *Computers in Human Behavior*, *23*(4), 1997–2010.

Acknowledgments