Expediting Support for Social Learning with Behavior Modeling

Yohan Jo[†], Gaurav Tomar[†], Oliver Ferschke[†], Carolyn P. Rosé[†], Dragan Gašević[‡]
[†]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
{yohanj, gtomar, ferschke, cprose}@cs.cmu.edu
[‡]Schools of Education and Informatics, The University of Edinburgh, Edinburgh, UK
dgasevic@acm.org

ABSTRACT
An important research problem for Educational Data Mining is to expedite the cycle of data leading to the analysis of student learning processes and the improvement of support for those processes. For this goal in the context of social interaction in learning, we propose a three-part pipeline that includes data infrastructure, learning process analysis with behavior modeling, and intervention for support. We also describe an application of the pipeline to data from a social learning platform to investigate appropriate goal-setting behavior as a qualification of role models. Students following appropriate goal setters persisted longer in the course, showed increased engagement in hands-on course activities, and were more likely to review previously covered materials as they continued through the course. To foster this beneficial social interaction among students, we propose a social recommender system and show potential for assisting students in interacting with qualified goal setters as role models. We discuss how this generalizable pipeline can be adapted for other support needs in online learning settings.

1. INTRODUCTION
More and more recent work in educational data mining and learning analytics refers to a "virtuous cycle" of data leading to insight on what students need and then improvements in support for learning [17]. An important goal is tightening this cycle to improve learning experience. We are interested especially in social learning, drawing from a Vygotskian theoretical frame where learning practices begin within a social space and become internalized through social interaction. This may involve limited interaction, such as observation, or more intensive interaction through feedback, help exchange, sharing of resources, and discussion.

There are two main contributions of this paper. The first is to propose a pipeline that can expedite the cycle of data infrastructure, learning process analysis, and intervention (Figure 1). Data infrastructure provides a uniform interface for heterogeneous data from social interaction in various platforms, such as connectivist Massive Open Online Courses (cMOOCs) [15], hobby communities, and Reddit communities, where people engage in follower-followee relations, post updates to their account, engage in threaded discussions, and also optionally link in blogs, YouTube videos, and other websites. Learning process analysis aims to analyze students' processes depending on their social network configurations and to identify beneficial kinds of social connections. We developed a probabilistic graphical model that analyzes sequences of behaviors in terms of topics expressed and social media types that students actively engage in over time. Finally, intervention is introduced to foster beneficial social connections among students. We developed a recommender system that matches qualified students to discussions to increase opportunities for them to interact with other peers. The pipeline is iterative such that data from participation is used to create models that trigger interventions in subsequent runs of the course. Data from those later runs can be used to train new and better models in order to improve the interventions, and so on.

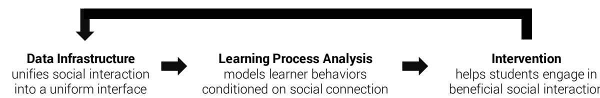


Figure 1: Pipeline for educational data mining in social learning.

Our second contribution is to present findings from an application of the proposed pipeline to data from a social learning environment called ProSolo [12], in order to investigate the positive influence of observing goal-setting behavior. While goal-setting has been intensively researched and proven to be an important self-regulated learning (SRL) practice that often leads to success in learning, the influence of a student's goal-setting behavior on observers has little been investigated empirically. If goal-setting students turn out to be good role models, that is, beneficial to their social peers, we can encourage and help students to make such social connections with goal setters to enhance their learning experience. The usefulness of this effect may be especially desirable in online courses where the number of instructors is limited, or online communities that are not structured like courses, where students are required to take more agency in forging a learning path for themselves within an ecology of resources.

In the remainder of this paper, we first motivate the specifics of our pipeline as situated within the literature. Next, we present our pipeline and its application, along with findings.

2. RELATED WORK

Vygotsky's view of social interaction as a key to learning and Bandura's social learning theory [1] emphasize the importance of interaction to learning. In social contexts, by vicarious learning, students observe external models and learn from those observations even when not actively engaged in interaction [19]. Observation of role models facilitates motivation and self-efficacy for a task [14] and may be associated with positive changes in the observer's behavior [9]. Drawing on this theoretical foundation, the positive impact of social interaction has been investigated in collaborative work [8] and in online courses [11]. Yet, to our knowledge, our work is the first to investigate goal-setting behavior specifically as a qualification of a role model in online learning.

Several data infrastructures have been introduced to aid educational data mining for Massive Open Online Courses (MOOCs). For instance, MOOCdb [18] and DataStage[1], designed to store raw data from MOOCs, consolidate clickstream data from different MOOC platforms in a single, standardized database schema. This allows for developing platform-independent analysis tools, thus enabling analyses that span multiple courses hosted by different MOOC providers with reduced development effort. While these infrastructures focus on behavior data represented by clickstream logs, our proposed infrastructure deeply represents other aspects of student interactions, such as discussion behavior and social relationships, which require the natural language exchange between students.

Analysis of students' learning processes has been a critical topic in education. Our method contributes to the literature on time series behavior modeling. Approaches to learning process analysis differ in the definition of the basic building block, often conceived of as states within a graph. Common building blocks for tutoring systems and educational games include knowledge components [22] and actions [13]. In dialogue settings, it is common to code each utterance according to a coding scheme and analyze the sequence of codes [4]. In a MOOC context, states are often defined as course units [3], course materials [3], or discussions [2]. Such predefined states, however, may not be the ideal units of states, especially in online courses where students can selectively engage in learning resources. Therefore, unsupervised modeling approaches are appealing for the purpose of identifying states that are meaningful indications of student interests obtained in a data-driven way. Our model belongs to the class of Markov models, which have been proposed to learn latent states and state transitions [6, 21].

In MOOCs, a student's learning process is affected by other peers especially through interaction in forums, which offer opportunities to develop communication and community. Hence, social recommendation algorithms can introduce appropriate students to certain discussions for productive interaction. Suggested matches should be appropriate when viewed either from the discussion or student side [16], for example by suggesting a student to participate in discussions based on both the potential benefit of the student's expertise as an asset to the discussions while respecting the limitations of a student's resources for participation in more than a limited number of discussions [20]. Our model can recommend discussions to a student by balancing the benefit of the student's qualification to discussions, her relevance to discussions, and required effort.

[1] http://datastage.stanford.edu/

3. THREE-PART ANALYTICS PIPELINE

Our pipeline is designed to expedite the process of exploiting student data leading to data-driven decision-making for enhancing student learning (Figure 1).

In this pipeline for social learning, the first component is a data infrastructure that maps diverse forms of social interaction into a common structure. This uniform interface allows the subsequent components—learning process analysis and intervention—to apply the same tools to different data, even from distinctly different discourse types, with little modification. Our development of this infrastructure, DiscourseDB[2], represents discourse-centered social interaction as an entity-relation model. Discourses (e.g., forums or social media) and individual contributions in a discourse (e.g., posts, comments, and utterances) are represented as generic containers generalizable to diverse social platforms. DiscourseDB also allows for defining arbitrary relations between contributions, e.g., a "reply-to" relation derived from the explicit reply structure of the platform versus one inferred through some automated analysis process. This flexibility helps the subsequent components of the pipeline avoid data-specific processing. DiscourseDB can store both active and passive activities of individuals, such as creating, revising, accessing, and following contributions, as well as forming social connections with other individuals. DiscourseDB is the key component of our pipeline, based on which the next components perform integrated analyses of discourses and social networking on multiple platforms with reusability.

The second component of our pipeline is analysis of students' learning processes depending on their social connections. The goal is to assess students' needs of support by understanding how learning processes are affected by social interaction and what types of social interactions are helpful to students. Just as Bayesian knowledge tracing enables modeling the learning process from a cognitive perspective and then supporting a student's progress through a curriculum, Bayesian approaches can model learning processes at other levels, including supportive social processes. And similarly, these models can then be used to trigger support for the learning processes in productive ways. Hence, the third component of our pipeline draws upon insights obtained from the analysis to introduce interventions that can help students make beneficial social connections with other peers. We will propose two concrete examples of machine learning techniques for these two components in Section 5 and Section 6 respectively.

4. APPLICATION OF PIPELINE

The remainder of the paper presents an example application of our general pipeline to a specific problem. We propose ex-

[2] http://discoursedb.github.io

ample models for learning process analysis and intervention that can build upon DiscourseDB. After this description we discuss our findings. This section introduces the data set for that exploration.

4.1 Problem and Data

We examine goal-setting behavior as a potential qualification of good role models via learning process analysis and foster social connections with goal setters via recommendation support. Since most MOOCs and informal learning communities lack a measure to identify potentially good role models (e.g., a pretest), increased frequency of effective goal-setting behaviors may serve as an indirect indicator of success, as previous studies showed positive relationships between goal-setting behavior and learning outcomes [5, 23].

The data was collected from an edX MOOC entitled *Data, Analytics, and Learning* (DALMOOC) [12], which ran from October to December 2014. This course covered theoretical principles about learning analytics as well as tutorials on social network analysis, text mining, and data visualization. This MOOC was termed a *dual layer* MOOC because students had the option of choosing a more standard path through the course within the edX platform or to follow a more self-regulated and social path in an external environment called ProSolo. The ProSolo layer allowed students to set their own learning goals and follow other students so that they could view activities and documents that offered clues about how to approach the course productively. While a huge literature on analysis of MOOC data focuses on Coursera, edX, and Udacity MOOCs, other platforms with more social affordances are growing in popularity. In order to serve the goal of identifying support needs and automating support that may be triggered in a social context, it is advantageous to work with data from socially-oriented platforms. We used the log data from ProSolo as our object of analysis, which include students' discussions on ProSolo and their own blogs and Twitter that they identified on their ProSolo profile pages, evidence of students' social connection with each other, and "goal notes," which students can use to set their learning goals in their own words.

We preprocessed discussion data before running our model. First, we filtered course-relevant tweets using the hashtags #prosolo, #dalmooc, and #learninganalytics. We confirmed that the tweets identified as irrelevant by this process have little to do with course activity. Because we are not interested in irrelevant content, we replaced such content with a tag to indicate irrelevant content. In order to prevent topics from being defined in terms of document types, we removed Twitter mentions and "RT" from tweets as well as other function words including URLs from all documents. Descriptive statistics for the data set are listed in Table 1.

Goal notes	62	Tweets (relevant)	715
ProSolo posts	318	Tweets (irrelevant)	25,461
Blog posts	359		
Users	1,729	Social connections	814

Table 1: Descriptive statistics for ProSolo data.

4.2 Goal Quality and Social Connection

To categorize the quality of goal-setting behavior of each student, we first annotated each goal note written by students indicating whether it indeed contains a goal or not. 58% of goal notes contained goals. An example goal note is as follows: *"to understand learning analytics and see how these may be useful for my teaching and in particular, my learning resource design/development."* On the basis of this annotation, we categorized students into three classes: (1) goal setters, (2) goal participants, and (3) goal bystanders. Goal setters have goal notes that mention their distal or/and proximal goals. Goal participants have goal notes, all of which are about something other than goals, e.g., experiences or questions. Goal bystanders have no goal notes. Note that the category of a student can change over time. All students start as goal bystanders and may become a goal participant or a goal setter as time passes. A student's *social connection* is then categorized into seven classes: (S1) has already been following a goal setter, (S2) started to follow a goal setter at the current time point (S3) has been following a goal participant (but no goal setter), (S4) started to follow a goal participant at the current time point, (S5) has been following a goal bystander (at best), (S6) started to follow a goal bystander at the current time point, and (S7) follows no one. S2, S4, and S6 mean that a student's social connection improved at the current time point, whereas S1, S3, and S5 indicate that a student remained in the same social connection category as in the previous time point.

5. LEARNING PROCESS ANALYSIS

Learning process analysis aims to assess students' needs of support. Hence, we model students' behavior and analyze their learning processes as they experience changes in their social connections throughout the course.

5.1 Model

Our model automatically extracts a representation of students' learning processes based on their discussions in a course and their social connections, which may reveal the influence of different configurations within the social space (see our technical report [7] for details). We define the building blocks of learning processes, i.e., states, in terms of discussed topics and the document types used for discussions (e.g. Twitter, blog). Given the sequences of timestamped documents and social connection types for students, our latent Markov model infers a set of states, along with the main topics and document types for each state. The learned topics reflect students' interests, and the document types show how students use different media for different interests. The model also learns transition probabilities between states, conditioned on the social connection category in the source state. This discloses how learning processes differ depending on students' social connection types.

5.2 Findings

We applied the model to the ProSolo data and examined the correlation between the categories of social connection and learning behaviors. We ran our model with the number of states set to 10 and the number of topics set to 20. We defined the unit of a time point as one week, and if a student had no activity in a certain week, that week was omitted from her sequence.

State	Topics	RelGoalNote	IrGoalNote	Post	Blog	RelTweet	IrTweet
0	Course-irrelevant tweets	0.00	0.00	0.00	0.00	0.00	1.00
1	Concept map, network analysis (Week 9)	0.00	0.00	0.02	0.01	0.18	0.78
2	Social capital (Week 3)	0.04	0.01	0.19	0.30	0.18	0.27
3	Tableau (Week 2), Gephi (Week 3), Lightside (Week 7)	0.01	0.03	0.10	0.28	0.24	0.34
4	Prediction models (Week 5)	0.01	0.02	0.29	0.22	0.10	0.36
5	Data wrangling (Week 2)	0.01	0.01	0.12	0.08	0.26	0.52
6	Visualization (Week 3)	0.05	0.02	0.24	0.47	0.08	0.15
7	Epistemology, assessment, pedagogy (Week 4)	0.05	0.00	0.18	0.22	0.30	0.25
8	Prediction, decision trees (Week 5)	0.02	0.02	0.19	0.40	0.09	0.28
9	Share, creativity (mixed topics)	0.00	0.02	0.12	0.13	0.21	0.52

Table 2: Learned states with their topics and document type distribution (each row sums to 1). (RelGoalNote: goal notes containing a goal, IrGoalNote: goal notes without a goal, Post: posts on ProSolo, Blog: personal blog posts, RelTweet: course-relevant tweets, IrTweet: course-irrelevant tweets)

	Social Connection			
	GS S1+S2	GP S3+S4	GB S5+S6	NO S7
# Time Points	139	315	265	821
% Time Points				
State 0	**$0.59^{\star\star}$**	**0.75**	**0.75**	**0.71**
State 1	**$0.17^{\star}$**	**0.10**	**0.03**	**0.04**
State 2	0.05	0.02	0.02	0.04
State 3	**$0.04^{\star}$**	**0.00**	0.01	**0.01**
State 4	0.01	0.02	0.03	0.06
State 5	0.05	0.03	0.06	0.05
State 6	0.05	0.02	0.02	0.02
State 7	0.03	0.01	0.03	0.02
State 8	0.00	0.03	0.02	0.02
State 9	0.01	0.04	0.03	0.04

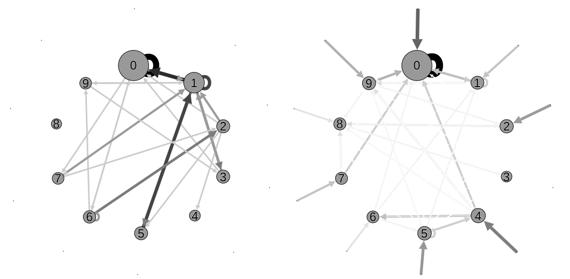
Table 3: Proportion of time points students stay in each state depending on the social connection (each column sums to 1). "$\star\star$" and "$\star$" indicate that GS is significantly different from other categories in bold with $p < 0.01$ and $p < 0.05$, respectively, by Pearson's χ^2 test. GS, GP, and GB each represent either "has been following" or "started to follow" a goal setter, a goal participant, and a goal bystander, respectively. NO means to follow no one.

5.2.1 *Learned States*

The model learns states with their topics and document type distributions (Table 2). Most states are aligned well with course units covering important course topics. However, State 0 is where students do not participate in course discussion but post course-irrelevant tweets. State 3 is about hands-on practice of software tools across the course, and State 9 covers many side topics. Tweets tend to take a large proportion and goal notes a small proportion in every state due to their relative volumes. Blog posts are actively used for summarizing readings and tutorials, and tweets are used as a means of communicating with lecturers (e.g., State 5). ProSolo posts are most accessible to ProSolo users, so students use them to reveal their opinions and questions.

5.2.2 *Students Following Goal Setters*

According to the investigation of students' learning processes, based on the number of weeks they spent in each state (Table 3) and state transition patterns (Figure 2), students who follow goal setters show the following positive learning behavior:



(a) S1. Following a goal setter (b) S7. Follows no one

Figure 2: State transition patterns. Nodes are states whose size reflects the number of weeks students visit the states. Edges are transitions whose thickness and darkness reflect transition frequency. Edges without a source node represent the probability of being the first state in a learning path.

Twitter usage: The students following goal setters spend noticeably fewer weeks on irrelevant tweets (State 0).

Participation duration: The topics of the states in which students stay reveal how long they persist in the course. The students following goal setters are more likely to discuss the material taught in the last week (State 1), that is, they are active in the last phase of the course.

Activities of interest: The number of weeks students spend in each state reflects the activities students are interested in. The students following goal setters were more active in hands-on practice (State 3) than other students. Hands-on practice requires higher motivation than merely watching lectures, so these students might have been helped by observation of role models as discussed in the literature [14]. This trend would have not been as clear using predefined states based on course units [3].

Study habits or challenges: Transition patterns may reveal students' study habits or challenges. Figure 2a shows frequent transitions between three states (States 1, 3, and 5) that are associated with materials taught in different weeks. Such transitions may reflect the SRL strategy of activating and applying prior knowledge to the current situation [10].

These positive effects associated with following goal setters are not apparent with other social connection types, e.g., following no one (Figure 2b). This indicates that "who to follow" is more important than simply following someone.

6. INTERVENTION FOR SUPPORT

On the basis of the insights obtained from the previous component, the third component of our pipeline is to offer appropriate support, especially towards fostering beneficial social connections between students. We argue that a recommender system can serve this purpose, by presenting its potential positive impact as assessed on the corpus.

6.1 Model

Our recommender system aims to match qualified students (e.g., goal setters) to discussions so that they can interact with and benefit the discussants through discussions (see our technical report [7] for details). Our model has two steps: relevance prediction and constraint filtering. The relevance prediction step learns the relevance between students and discussions using student- and discussion-related features. The learned relevance reflects students' preferences and tendencies, but may not reflect the ideal matches for fostering learning. The constraint filtering step thus combines the relevance scores with some constraints that foster interaction between qualified students and other students, and finalizes recommendations.

6.2 Findings

Since we have identified positive learning behaviors of students who follow goal setters, we may want to support students by fostering interaction with goal setters. Instead of recommending direct following relations, which are not supported by many learning platforms, we recommend discussions to qualified students so that they can interact with the discussants. We first assess the extent to which students are sensitive to qualified students prior to explicit intervention, and then present the potential added value of our recommendation model.

6.2.1 *Students' Awareness of Role Models*

Our first step is to assess whether students can identify effective role models in discussion activities (ProSolo posts), by measuring the impact of the information about students' qualifications on the prediction of discussion participation. This task is to infer links between students and discussions that we hid from an observed static snapshot of a network of discussion participation based on observable data. A measured positive impact here would indicate some sensitivity on the part of students to interact with qualified students naturally. We train a predictive model of students' participation in discussions on two thirds of student-discussion pairs. We then predict the discussion participation of the remaining pairs. Our evaluation metric is mean average precision (MAP).

We compared four configurations by varying the information about students' qualifications that is used as feature for relevance prediction. In particular, CAMF uses only basic features, such as the numbers of discussions each student initiated and participated in and each discussion's length, number of replies, and participants. CAMF_G and CAMF_C add information about goal quality and degree centrality, respectively, and CAMF_GC adds both. The evaluation was conducted as a link prediction task, based on the relevance scores predicted in the relevance prediction step. Students' qualification information did not improve link prediction accuracy (Table 4). This means that students are not proactively sensitive to peers' qualifications while participating in discussions, which supports our view that explicit recommendation could be valuable for encouraging students to interact with qualified peers through discussions.

Configuration	MAP	Configuration	MAP
CAMF	0.465	CAMF_C	0.455
CAMF_G	0.438	CAMF_GC	0.439

Table 4: MAP for link prediction.

Configuration	OB	Configuration	OB
GoalPart	1.888	MCCF_G	3.683
HighCent	1.943	MCCF_C	3.770
GoalPart_HighCent	1.873	MCCF_GC	3.656

Table 5: Overall Community Benefit for recommendation.

6.2.2 *Recommendation Quality*

The recommendation of discussions should be consistent with both the relevance between students and discussions (the relevance prediction step) and constraints for beneficial social connection (the constraint filtering step). To this end, we evaluated recommendation quality on Overall Community Benefit (OB) [7]: the relevance of our recommendations penalized by the burden on the students induced by the recommendations. The higher OB the better.

We tested three configurations by varying the constraints incorporated into the constraint filtering step. MCCF_G requires that every discussion have at least one goal participant or goal setter. MCCF_C requires that every discussion have at least one student whose degree centrality is higher than 0.1. MCCF_GC requires both. In addition, the following configurations were tested as baseline without incorporation into the model. GoalPart filters goal participants or goal setters after making recommendations based on predicted relevance. Similarly, HighCent filters students with degree centrality higher than 0.1. GoalPart_HighCent filters goal participants or goal setters with degree centrality higher than 0.1. Incorporating the constraints about students' goal quality and degree centrality into the model (MCCF_G, MCCF_C, and MCCF_GC) achieved higher OB than the simple filtering approaches (Table 5). That is, our algorithm effectively matches qualified models to relevant discussions in such a way that students in every discussion can interact with qualified models while balancing the load of the models.

7. DISCUSSION

According to our learning process analysis, students benefit from social connections with effective goal setters through ProSolo's follower-followee functionality. They stay longer in the course, engage in hands-on practices, and link materials across the course. This supports the view that goal-setting behavior is a useful qualification for potential role models. According to the discussion participation prediction task, explicit intervention is important for helping students be aware of qualified students and interact with them via discussions. Therefore, we incorporated the information about students' qualifications into our recommendation model as constraints, successfully matching qualified learning partners to relevant discussions.

This work started from the need for expediting data analysis and analysis-informed support in social learning where students interact with one another via various social media in order to pursue their own learning goals. This expedition builds on DiscourseDB, data infrastructure for complex interaction data from heterogeneous platforms. We proposed a probabilistic graphical model to analyze students' learning processes depending on the state of their social connections, and proposed a recommender system that can improve student support on the basis of the insights obtained from the analysis. This pipeline arguably should allow us to apply the techniques to different learning communities with little effort.

Goal-setting behavior is an important practice in SRL and is known to be difficult for students, so an analysis towards improvement of this skill is arguably valuable. Nevertheless, in this study we have not examined how this behavior influences the domain learning of students. This is due both to the limited data size for our first trial to use ProSolo in MOOCs as well as a lack of learning gain measures. However, the modeling techniques proposed in this paper can readily be applied to other data sets if the requisite data become available. We are also interested in investigating different SRL strategies besides goal-setting in social learning, and how social interaction influences the SRL behaviors of the students. Ultimately, the real value of the work will be demonstrated not with a corpus analysis, as for our proposed recommendation approach, but with an intervention study in a real MOOC. We are working towards incorporating this approach in a planned rerun of DALMOOC.

8. ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under grants ACI-1443068 and IIS-1320064, and by the Naval Research Laboratory and Google.

9. REFERENCES

[1] A. Bandura. *Social Learning Theory*. Morristown, N. J.: General Learning Press, 1971.

[2] A. Bogarín and R. Cerezo. Discovering students' navigation paths in Moodle. In *EDM '15*, pages 556–557, 2015.

[3] C. Coffrin, L. Corrin, P. de Barba, and G. Kennedy. Visualizing patterns of student engagement and performance in MOOCs. In *LAK '14*, pages 83–92, Mar. 2014.

[4] A. Ezen-Can, K. E. Boyer, S. Kellogg, and S. Booth. Unsupervised modeling for understanding MOOC discussion forums. *LAK '15*, pages 146–150, 2015.

[5] J. Husman and D. F. Shell. Beliefs and perceptions about the future: A measurement of future time perspective. *Learning and Individual Differences*, 18(2):166–175, Apr. 2008.

[6] Y. Jo and C. P. Rosé. Time Series Analysis of Nursing Notes for Mortality Prediction via a State Transition Topic Model. In *CIKM '15*, 2015.

[7] Y. Jo, G. Tomar, O. Ferschke, C. P. Rosé, and D. Gašević. Expediting support for social learning with behavior modeling. *arXiv:1605.02836*, 2016.

[8] I. Molenaar and M. M. Chiu. Effects of sequences of socially regulated learning on group performance. In *LAK '15*, pages 236–240, Mar. 2015.

[9] E. L. Paluck, H. Shepherd, and P. M. Aronow. Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, 113(3):566–571, Jan. 2016.

[10] P. R. Pintrich. A Conceptual Framework for Assessing Motivation and Self-Regulated Learning in College Students. *Educational Psychology Review*, 16(4):385–407, Dec. 2004.

[11] C. P. Rosé, R. Carlson, D. Yang, M. Wen, L. Resnick, P. Goldman, and J. Sherer. Social factors that contribute to attrition in MOOCs. In *L@S '14*, pages 197–198, Mar. 2014.

[12] C. P. Rosé, O. Ferschke, G. Tomar, D. Yang, I. Howley, V. Aleven, G. Siemens, M. Crosslin, D. Gasevic, and R. Baker. Challenges and Opportunities of Dual-Layer MOOCs: Reflections from an edX Deployment Study. In *CSCL '15*, pages 848–851, 2015.

[13] E. Rowe, R. S. Baker, and J. Asbell-Clarke. Strategic game moves mediate implicit science learning. In *EDM '15*, pages 432–436, 2015.

[14] D. H. Schunk and A. R. Hanson. Peer models: Influence on children's self-efficacy and achievement. *Journal of educational psychology*, 77(3):313–322, 1985.

[15] G. Siemens. Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, 2014.

[16] L. Terveen and D. W. McDonald. Social matching: A framework and research agenda. *ACM transactions on computer-human interaction*, 12(3):401–434, 2005.

[17] C. Thille. Education Technology as a Transformational Innovation. *White House Summit on Community Colleges: Conference Papers*, pages 73–78, 2010.

[18] K. Veeramachaneni, S. Halawa, F. Dernoncourt, U. O'Reilly, C. Taylor, and C. Do. Moocdb: Developing standards and systems to support MOOC data science. *CoRR*, abs/1406.2015, 2014.

[19] P. H. Winne and a. F. Hadwin. Self-regulated learning and socio-cognitive theory. *International Encyclopedia of Education*, pages 503–508, 2010.

[20] D. Yang, D. Adamson, and C. P. Rosé. Question recommendation with constraints for massive open online courses. In *RecSys '14*, pages 49–56, 2014.

[21] J. Yang, J. McAuley, J. Leskovec, P. LePendu, N. Shah, and B. Informatics. Finding Progression Stages in Time-evolving Event Sequences. In *WWW '14*, pages 783–793, Apr. 2014.

[22] C. Zhao and L. Wan. A shortest learning path selection algorithm in e-learning. *Int'l Conference on Advanced Learning Technologies*, pages 94–95, 2006.

[23] B. J. Zimmerman. Goal setting: A key proactive source of academic self-regulation. In *Motivation and self-regulated learning: Theory, research, and applications*, pages 267–295. Erlbaum, 2008.