Revealing Interaction Patterns Among Youth in an Online Social Learning Network Using Markov Chain Principles

Sarah Bishara, Jennifer Baltes, Taha Hamid, Taihua Li, Denise C. Nacu, Caitlin K. Martin, Jonathan Gemmell, Chris MacArthur, Daniela Raicu, and Nichole Pinkard
sbishara44@gmail.com, jenniferbaltes1@gmail.com, tahahamiid@gmail.com, taihua.ray.li@gmail.com, dnacu@depaul.edu, cmartin@digitalyouthnetwork.org, jgemmell@depaul.edu, chrismacarthur@mac.com, draicu@depaul.edu, npinkard@depaul.edu
DePaul University

Abstract: The problem of the digital divide has shifted attention from access to inequities of participation and opportunities to develop 21st century skills in online learning platforms. In this paper, we explore Markov chain principles in a time-based probabilistic graphical approach to analyze a multi-year data set of log data generated by students from one urban middle school and coded using a framework aligned with 21st century learning activities. Results showed the efficacy of applying Markov chain principles in helping reveal similar and distinct usage patterns of the learners in this community across different time spans. This work has implications for the design and analysis of online learning platforms and for creating opportunities to help youth build 21st century skills using online learning platforms.

Introduction

The notion of a "digital divide" has shifted from a focus on inequities of access to *equipment* toward one of inequities of access to *opportunities*. There is a recognized "opportunity gap" that characterizes differences in access to learning activities and networks, resulting in a "participation gap" between youth in underserved communities and their more affluent counterparts (Warschauer & Matuchniak, 2010; Watkins, 2011; Lenhart, 2015). Computational and technological learning experiences have been linked to the development of skills and dispositions that are viewed as critical for participation in the 21st century, such as communication, creative production, problem solving, and collaboration (Ito et al., 2009; Barron, Gomez, Pinkard & Martin, 2014). A concerning consequence of this existing gap, then, is that certain populations of youth have fewer opportunities to develop these skills necessary for productive life and citizenship in today's world (Levy & Murnane, 2012).

Online social learning networks have affordances for supporting 21st century learning in ways that can potentially bridge inequities by offering learning opportunities, resources, and social supports beyond the physical boundaries that demarcate underserved populations (AITF, 2014; Hamid, Waycott, Kurnia, & Chang, 2015; Jenkins, Purushotma, Weigel, Clinton, & Robinson, 2009). As teachers in formal and informal learning settings are increasingly willing to make use of such systems (MMS Education, 2012), and as schools and districts increasingly adopt and require the use of online platforms (Burch & Good, 2014), there is anticipation that the gap will be minimized. These systems generate massive use data, and there is excitement around the potential of harnessing this data to reveal insights about learning and to design interventions that can level the playing field. *Who is participating? What resources are being used? Who is supporting whom?* To make progress on understanding learning patterns in user logs generated from online social learning networks, collaborative work is necessary to bring together learning theory, deep contextual understanding, and mathematical algorithms (Bienkowski, Feng, & Means, 2012; Pea, 2013; Siemens, 2012).

This study joins learning sciences and data science methods to explore patterns in user trace log data from students in a public urban middle school using an online social learning network over multiple years. We used a sequential pattern data mining technique, Markov chains, to explore student actions as opportunities for building 21st century skills. We asked, *how can Markov chain principles be used to reveal patterns of online activity over time in the areas of creative production, social learning, and self-directed learning?* In addressing this question, we aim to inform research methodologies used to study online learning platforms and to identify use patterns that can be later explored for their potential to provide evidence of 21st century learning. While attention to 21st century skills is increasing, we don't yet have practical ways to measure them and to understand how experiences using online social learning networks may contribute to building those skills. To address the "participation gap," developing ways to reveal how youth may be using these systems differently is needed.

Markov chains work on the assumption that the probability of the next action taken is exclusively dependent on the current action. For this study, a Markov chain model represents each action as a node in a network graph and any existent relationship between two actions as an edge connecting the nodes corresponding to those actions. Probabilities are associated with each edge to encode the strength of the relationships based on the temporal log data over a period of 28 months.

Related work

We contextualize this work within ongoing efforts to design and study socio-technical systems that help youth develop skills and competencies by interacting with peers and adult educators around the creation, sharing, and communication of digital artifacts. We define these systems as online social learning networks, web-based environments that use features of social network sites and learning management systems to support and develop an online learning community and the individual participants within it (Martin, Nacu & Pinkard, 2016). Youth participation in online networks has been linked to fostering 21st century skills such as self-directed learning, creativity, and communication (Hamid, et al., 2015; Ito et al. 2009; Jenkins et al. 2009). A focus on these types of skills is becoming increasingly emphasized in K-12 standards striving to prepare youth for future workforce needs (Common Core State Standards Initiative, 2010; NGSS, 2013; Pellegrino & Hilton, 2013).

Inequities exist in terms of who has the opportunities to participate in programs and activities that can build technological competencies related to 21st century learning. While youth are increasingly participating in social network sites (Blair, Millard, & Woolard, 2015; Lenhart, 2015; Watkins, 2011), studies have revealed that contributors of online content, in general, are a small subset of the population using technical systems (Rideout, 2015), and that this subset is not representative of the larger population (Glott, Schmidt, & Ghosh, 2010). Youth from areas with fewer socioeconomic resources are especially underserved, demonstrating inequities in more sophisticated forms of participation such as interest-driven practices involving creating, sharing, communicating, and critiquing (Margolis, Estrella, Goode, Holme, & Nao, 2010; Warschauer & Matuchniak, 2010).

In the last five years, a growing body of research studies have applied data mining techniques to examine log data generated by online learning platforms. These data have been analyzed in the context of user-to-user social interaction (Cela, Sicilia, & Sánchez, 2015; Xu, 2011), user-to-platform interaction (Kardan, Roll, & Conati, 2014), and user-to-content learning behaviors (Jeong et al., 2008). Some of these studies focus their analysis on Markov model chains (Faucon, Kidzinski, Dillenbourg, 2016; Marques & Bello, 2011). Rodríguez and Boyer (2015) analyzed probabilistic chains to compare problem-solving approaches by individual versus collaborative users. Within the educational domain, some prior studies have applied Markov chains to improve intelligent tutoring systems while others have used students' answers to multiplication problems to generate Markov chains that informed a question recommendation system (Taragari, Saranti, Ebner, & Schön, 2014).

Research context

This work is part of a multi-year study examining interactions among youth and adults in online social learning networks. We used data from an urban middle school ELA (English Language Arts) teacher and his 54 students. The teacher remained their ELA teacher for three academic years. Students in this study were in 6th grade in the first year, and, by the third year, were 8th graders. The K-8 school draws the majority of students from a predominantly Latino community: 91% of students are Latino, 8% black, and 1% white. In the school, almost half of students (43.3%) are classified as having limited English (English Language Learners) and 95% are classified as coming from low income households. The student sample in this work reflected the larger school demographics, with 47.2% girls, 52.8% boys, 89% self-reporting as Latino, 6% black, 2% white, and 2% Chinese. Over three quarters of students in our sample (83%) reported being part of Spanish-speaking households.

The study involved iRemix, which has an interface and functionality similar to popular online social networks (Barron et al., 2014; Martin, Nacu & Pinkard, 2016; Zywica, Richards & Gomez, 2011). Students and teachers can share digital artifacts such as blog posts, photos, and videos. They can respond to posts using comments and reactions, browse activity through a feed, edit profile pages, and link to peers. iRemix is intended to support the development of 21st century skills through production, reflection, critique, and revision. While the system is meant to be youth-driven and reflects youth interests, teachers can post challenges to prompt activities.

Method

Log data

Logs of student actions were pulled for the period from 1/1/2014 through 4/15/2016, spanning three academic years. Each observation in the data generates a row with the student's user ID, the associated action code, the timestamp, and other information not covered in this paper (Nacu, Martin, Schutzenhofer & Pinkard, 2016).

The framework for coding student actions builds on prior work which conceptualizes the activities the system was designed to support as opportunities for 21st century online learning. Specifically, the learning opportunities on iRemix were categorized into three themes that reflect the platform's learning goals: *creative production*, *self-directed learning,* and *social learning* (Martin, Nacu & Pinkard, 2016). *Creative production* involves developing identity as a creator, creating media, and revising work. *Self-directed learning* involves using online resources, monitoring one's progress, and seeking support and learning opportunities. *Social learning* involves communicating with others and observing the work of peers. For example, sending a message to another user, commenting, and posting a reaction on another user's work are considered *communicate* actions, and relate to the theme of *social learning*. Table 1 shows the action codes present in the dataset (with abbreviations), mapped onto the three focal themes of 21st century learning opportunities. After extensive data cleaning and preparation, there were 32,895 rows of coded data covering 10,193 sessions.

Table 1: Student action codes logged from user trace activities on iRemix

Abbreviation	Action Code	Description
Creative Production		
per	personalize	Change user profile information or avatar picture
cre	create	Submit original work (media artifact)
sha	sharing	Share a media artifact
bw	begin work	Start a set of scaffolded learning activities
eow	edit own work	Edit a submitted artifact created by self
gb	get badge	Complete set of scaffolded learning activities
row	review own work	View artifact created by self
cw	complete work	Submit a final draft of project
Self-Directed Learning		
vr	view resource	View a resource associated with a learning activity
vpa	view potential activity	View a description of a potential learning activity or set of activities
Social Learning		
com	communicate	Send a message, leave a comment or reaction on work,or reply to forum
cri	critique	Provide star-ratings for submitted work
vpo	view profile of others	View a user profile of another user
vwo	view work of others	View work submitted by another user
jc	join community	Join an interest group
vc	view community	View content of interest group
inv	invite	Invite other users to an interest group
qc	quit community	Leave an interest group
Other		
lgi	login	Login to system
lgo	logout	Artificial action indicating end of session

Constructing Markov models

With sequential coded data, a transition matrix can be computed. Each entry, *T(i,j),* in the transition matrix represents the probability of the action A_i to the action A_j. For example, *T(row|eow) =0.36* shows that the percentage of times a student who has just completed a "review own work" (*row*) action will complete an edit own work (*eow*) action next is 0.36. Likewise, T(*per*|*per*) *=0.13* shows that a student who has just completed a *personalize* (*per*) action is likely to complete another personalize action 13 percent of the time. Four transition matrices were created for this study: one for the full dataset spanning all three years, and one for each of the three academic years. The size of each matrix was (20x20), representing the twenty action codes.

Building on transition matrices, we constructed Markov chain models. A Markov chain is a stochastic process where a future state prediction is dependent only on the present state and does not take into account any information about the previous states. Conceptually, the Markov chain can be perceived as users moving from one state to another, and it can be visualized by network graphs. Due to the abundance of action-to-action activity that occurred in the logs, we only include edges representing a probability of 1% or more to increase readability of the network graphs. The thickness of the node's outline corresponds to the frequency of the action, highlighting the more frequent items. The thickness of the edges correlates to the probability of the action-to-action movement.

Results

Here, we first discuss how the Markov models can be interpreted and present insights from a network graph aggregating students' online activities over the full 28 months. Then, a year-by-year analysis follows.

Multi-year Markov model

The nodes of this Markov model represent the action types (see Table 1). The connecting edges between the nodes indicate both the direction and magnitude of the relationship between actions. For example, at the bottom right side of Figure 1, an edge exists from *eow* to *row* located near the left side of the graph, representing the 99% probability of students moving from editing their own work to reviewing their own work. In this case, the strong correlation is due to the internal workflow of iRemix, which automatically shows a user the work they have just edited. In comparison, *row* connects to *eow* with a thinner line, indicating a lower probability (36%) of a user returning to edit their work after reviewing it. Though the model shows any probability greater than 1%, actions occurring with a probability 10% or greater are significant given the number of possible transitions.

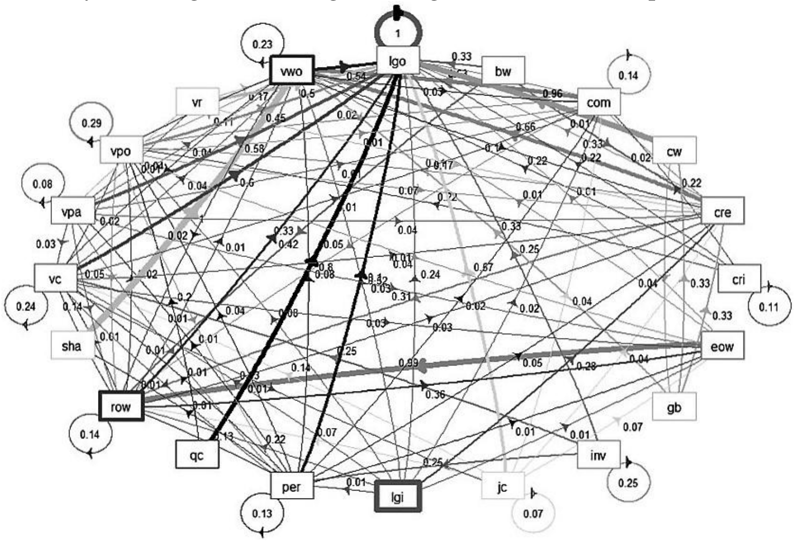

Figure 1: Multi-Year Network Graph.

Nodes can connect to themselves. For example, *row* transitions to itself (shown in the circle overlapping with the node) 14% of the time, indicating the likelihood of students reviewing their own work twice in a row. There are two unique states on the network graph: logging in (*lgi*) and logging out (*lgo*). *Lgi* is the beginning state of every user sequence. Therefore, no other nodes point toward it. *Lgo* is the absorbing state for the model; ultimately all users will end by logging out, thus no edges originate from it.

A key advantage of the Markov model is its capacity to describe the sum of actions taken over the entire program. Although there is an enormous amount of information captured by the graph, next we will focus on several of the larger patterns as they relate to our research questions.

Opportunities for *creative production*

One of the most notable transitions between nodes occurs between *row* and *eow*. In iRemix, editing actions include both altering the content of a previously submitted post as well as deleting the post. The *row* actions represent a student looking over their own post or looking at the assignment description. The system automatically directs students who have submitted work to review it, thus creating an extremely strong edge, 99% from *eow* to *row*. The opposite pathway is not automatic and occurs 36% of the time, suggesting students may be editing their work, viewing it, and then returning to the editing options. While this cyclical pattern of editing and reviewing would not be surprising to educators, this method helps to reveal that students are using the system in a way that is related to creative production of work, and reflects a type of desired activity. While not shown in the graph, this type of analysis can lead to further examination about which students are exhibiting this pattern and which are not. From an equity perspective, it is important to discern if and how students are participating online as creative producers.

Cre also frequently leads to *vwo* with a probability of 22%. This indicates a pattern of students posting their work then comparing their own to those posted by their peers. Qualitative studies have shown that looking at the work of others is often associated with generating new ideas and pushing new possibilities for the quality of submitted content (Ito et al., 2009; Barron et al., 2014).

Opportunities for *social learning*

One finding related to *social learning* is the connection between communicating and exploring the work of others in the community. Figure 2 shows that, across the 28 months, 62% of *com* actions end with a student transitioning to *vwo*. In iRemix, viewing others' work includes looking at another user's portfolio of work or individual posted artifact. Communicating can take many forms, including sending a message, responding to a user's work with a reaction, or contributing to a forum. This suggests that after a communication-related move, a student is very likely to keep exploring work submitted by others. This idea of an exploratory activity pattern is supported by the finding that 23% of *vwo* actions result in another *vwo* action. This pattern is encouraging as it reveals a social learning pattern in which communication actions lead to potential for learning by viewing work submitted by others. *Vwo* is also highly correlated to *gb* (33% probability) indicating that, after finishing a challenge, students are either looking for comparison pieces on the current challenge, or they are seeking inspiration for a new task.

In contrast, *vwo* to *com* does not have as strong of a connection in our data, as this chain of actions is only present for 3% of the pathways stemming from *vwo*. As a way to understand the potential for social learning observed in the data, we may interpret this finding to suggest that even when the work of others is viewed, users are not likely to provide feedback through comments or reactions, thereby missing learning opportunities for both the viewer and creator of posted work. While communication happened during the exploratory activity patterns, as evidenced above, communication actions, in general, were much less frequent in this dataset than other, more passive activity on the system. The Markov model easily revealed this disconnect in a way that has encouraged program educators and designers to think about ways to support youth to engage more in a communicative interaction around the work after viewing it.

One of the most telling sequences is between actions and themselves. These are referred to as "sticky states," actions which frequently lead to another instance of the same action. *Com*, *cri, inv*, *per, vc*, *vpo,* and *vwo* are all social actions. These are also states which have a self-referential probability of greater than 10%. This suggests that once engaged in social activity, users are likely to continue their social action. Although the connections between social actions are not substantial individually, edges originating from social actions are numerous. For example, *vpo* has a greater than 1% probability of leading to one of five social actions (including itself). These chains indicate that once social action has been initiated, it is likely to continue to another form.

Opportunities for *self-directed learning*

One goal for iRemix is to foster a space in which students have the social support and access to resources that allow students to take initiative in pursuing interests. In this dataset, we observed two actions aligned with this 21st century learning theme of *self-directed learning*: *vr* and *vpa*. We found that *vr* led to *cre*, *vpa* or *vwo* 16.7% of the time for these actions, suggesting that after viewing a resource associated with a learning activity, students submitted a work artifact, viewed a description of a potential learning activity or set of activities, or looked at others' work. Encouragingly, this finding highlights connections among actions cutting across the three 21st century learning themes. These observed connections among using resources, creating artifacts, viewing the work of peers, and exploring what challenges to take on next would indicate a sophisticated form of participation in the online system that is desired over other, more passive forms.

Year-by-year Markov models

The Markov models in Figure 2 each represent a single year of sequential data and provide the opportunity to evaluate how activity on iRemix changed over time. For Year 1, the system was adopted in the winter of the first year and thus accounts for approximately 6 months of the school year. For Year 3, data was pulled at the end of the winter; thus, the third year has approximately 4 months' worth of activity. However, as Markov models deal with transition probabilities and not counts, the models can still be compared with one another.

Many of the same observations from the multi-year Markov model (see Figure 1) are also evident in the yearly models. A strong connection is still observed between *eow* and *row*. Likewise, *cre* actions often transition to *vpo* across all three years, as they do in the multi-year model.

One apparent difference in use patterns of students over time is that by year three there were fewer actions present in the model. For example, *per* does not occur at all in the third year. An action is coded as *personalize* when a user updates information in his/her profile or posts a status to be shared with the community. This pattern can be explained by the notion that users would be inclined to spend more time personalizing their profile as a community is first developing. By the eighth grade in Year 3, it is likely that personas are more solidly established through face-to-face and online interactions, reducing the need for updating online profiles. From an educator's point of view, being able to discern changes in personalization may trigger intervention in which students may be encouraged to update their profiles to reflect their changing interests and identities within the learning community.

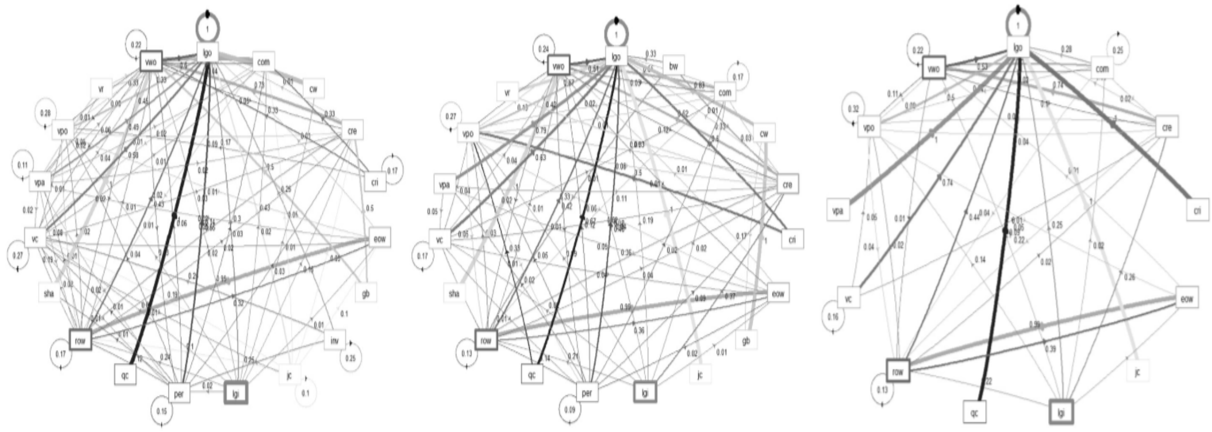

Figure 2. Network graphs by academic year: Year 1, Year 2, Year 3 (left to right).

Activities reflective of *social learning* such as *vwo*, *vpo,* and *vc* played out similarly, whereas *inv*, an action logged when a user invited another user to a community, appears in the Year 1 graph but not in the subsequent year graphs. This may indicate the developmental change of the community; creating and participating in interest groups may have had novelty and momentum in Year 1, but later dropped off as the community evolved. Again, this suggests an opportunity for the teacher to encourage students to engage in interest-based activities.

The social trends visible in the network graph evolved over the course of the program. In Year 1, 97% of *com* actions led to *vwo*. In Year 2, students performed a wider variety of actions, including *cre*, *row*, *vpo,* and *lgo*. *Vwo* remained the most likely transition at 54%, while another *com* action occurred 17% of the time. By Year 3, *vwo* was likely in 37.5% of transactions, while the self-referential probability increased to 25%. *Cre, row, vpo,* and *lgo* were also visible, with *lgo* having a probability of 27%. These findings add another dimension to the patterns of social learning visible in the multi-year Markov model; although social actions frequently follow one another, the learning community used communicate actions differently across the three years.

Discussion

The purpose of this study was to explore how Markov chain principles could be used to reveal patterns of students' online activity over time in the areas of *creative production*, *social learning*, and *self-directed learning*. Using a coding framework and Markov chain principles to discern frequent sequences of actions, this approach was useful for visualizing students' use of platform features with respect to goals to foster 21st century learning activities. We explored this approach with the notion that being able to understand how youth are making use of online social learning platforms is needed in order to attend to equity in terms of the quality of participation. Indeed, by analyzing at the level of a user session (rather than by single action, or by week, for example), the network graphs efficiently captured episodes of student online interaction in terms of types of desired activities.

In the three-year Markov model, we observed cyclical movement between editing and viewing own work, indicating the theme of *creative production*. We also saw a connection between *creative production* and *social learning* through the relationship between creation actions and viewing work own work and the work of others. *Social learning* occurred frequently with *creative production*, suggesting an underlying feedback loop between these two types of activities. Taken as a whole, the findings revealed by the Markov graphs show that students using iRemix were engaged in activity that was desired by the teacher: creative production characterized by iteration and supported by social interactions among the community. While the results do not present definitive evidence that learning actually occurred, they do demonstrate that the online learning platform provided opportunities to build 21st century learning skills for students (Reich, Murnane & Willett, 2012). The methodology presented here suggests one way to reveal those opportunities.

The network graphs produced for each year provided a concise way to examine how this particular learning community changed over the three year period. In particular, the graphs helped to show the continuity of core activities that were prompted and encouraged by the teacher including the posting of work, revising, and communicating, as well as types of actions that tended to drop off over time. For designers, this type of analysis can help generate insights that can inform how desired learning cycles can be strengthened by making the connections between these activities more obvious to students. The probabilities generated by these models have revealed existing patterns present in student work cycles, but providing personalized prompts to users could increase the likelihood of target behavior. Seeing how features are used or not over time may also suggest the need for teacher prompting to use desired features or may point to potential usability or logging issues that need to be further investigated. In terms of implications for developing theory, examining differences across time can give researchers a developmental view of a changing learning community, and the evolution of how the teacher used iRemix as part of his class over the three academic years. More research analyzing longitudinal log data using this type of approach is needed to compare high-level patterns which may provide a view on how a community of learners evolves and matures. While not addressed in this paper, our qualitative data (e.g., field observations, interviews, and surveys) can aid interpretation of these long-term trends.

The method explored here resulted in graphs that could be used by researchers to study patterns of use; however, such visualizations are likely not usable for educators. Additional research and design with educators is needed to better communicate insights in a way that integrates with practice. Furthermore, although the network graph is useful for determining global, longitudinal patterns, it does not provide information about individual students' use in a way that is helpful to the classroom teacher. The current visualizations, then, represent an initial step to reveal patterns which can establish a baseline. Future applications will allow identification of groups that are distinguished by their use patterns in order to provide the teacher with recommendations for action.

One of the main limitations of log data analysis is accurately interpreting user intent. Although some of the interpretations may be accurate for a group of students, further investigation is required to more deeply understand why students are exhibiting certain sequences of actions in the online platform. A key implication of this work is the need for diverse research collaborations to connect an understanding of intent with analytic methods used. Educators, learning scientists, data analysts, and software designers are only a subset of the types of researchers whose insights will provide a well-rounded interpretation of these types of data.

Conclusions

In this paper, we examined how Markov model principles could be applied to student action log data from the iRemix platform to visualize and better understand online behavior patterns with respect to 21st century learning opportunities. Network graphs generated based on the proposed Markov chain approach quantified the strengths of any existent pair of actions over different time spans, revealing distinctions and similarities in online behavior data. This approach is encouraging, as it shows potential for generating insights that can be useful for educators, designers, and researchers. Particularly for efforts aimed at addressing the "participation gap" characterized by inequities in terms of the development of 21st century skills among youth, this method suggests an approach that can reveal the nature of online interactions in a holistic and concise manner. Future work will explore new visualization techniques for predictive models of students' online interactions using Markov principles.

References

AITF on Learning and the Internet. (2014). *Learner at the center of a networked world.* Washington, DC: The Aspen Institute.

Barron, B., Gomez, K., Pinkard, N., & Martin, C. K. (2014). Digital Youth Network: Cultivating digital media citizenship in urban communities. Cambridge, MA: MIT Press.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *US Dept. of Education, Office of Educational Technology,* 1-57.

Blair, R., Millard, D., & Woollard, J. (2015). Perceptions of school children of using social media for learning. *International Journal on E-Learning*.

Burch, P., & Good, A. G. (2014). *Equal scrutiny: Privatization and accountability in digital education.* Harvard Education Press.

Cela, K. L., Sicilia, M. Á., & Sánchez, S. (2015). Social network analysis in e-learning environments: A preliminary systematic review. *Educational Psychology Review*, 27(1), 219-246.

Common Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics*. Wash., DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Faucon, L., Kidzinski, L., & Dillenbourg, P. (2016). Semi-Markov model for simulating MOOC students. In Barnes, T., Chi, M., & Feng, M. (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining*. Paper presented at the 9th International Conference on Educational Data Mining.

Glott, R., Schmidt, P., & Ghosh, R. (2010). *Wikipedia survey–overview of results.* United Nations University: Collaborative Creativity Group.

Hamid, S., Waycott, J., Kurnia, S., & Chang, S. (2015). Understanding students' perceptions of the benefits of online social networking use for teaching and learning. *The Internet and Higher Education*, 26, 1-9.

Ito, M., Baumer, S., Bittanti, M., Cody, R., Stephenson, B. H., Horst, H. A., Perkel, D., et al. (2009). *Hanging out, messing around, and geeking out: Kids living and learning with new media.* MIT press.

Jenkins, H., Purushotma, R., Weigel, M., Clinton, K., & Robison, A. J. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century.* MIT Press.

Jeong, H., Gupta, A., Roscoe, R., Wagster, J., Biswas, G., & Schwartz, D. (2008, June). Using hidden Markov models to characterize student behaviors in learning-by-teaching environments. In *International Conference on Intelligent Tutoring Systems* (pp. 614-625). Springer Berlin Heidelberg.

Kardan, S., Roll, I., & Conati, C. (2014, June). The usefulness of log based clustering in a complex simulation environment. In Trausan-Matu, S., Boyer, K., Crosby, M., Panourgia, K. (Eds.), *Intelligent Tutoring Systems*. Paper presented at the 12th International Conference on Intelligent Tutoring Systems (pp. 168-177). Springer International Publishing.

Lenhart, A. (2015). *Teens, social media & technology overview 2015.* Washington, DC.

Levy, F., & Murnane, R. J. (2012). *The new division of labor: How computers are creating the next job market.* Princeton University Press.

Margolis, J., Estrella, R., Goode, J., Holme, J. J., & Nao, K. (2010). *Stuck in the shallow end: Education, race, and computing.* Cambridge, MA: MIT Press.

Marques, A., & Belo, O. (2011). *Discovering student web usage profiles using Markov chains.* Electronic Journal of e-Learning, 9(1), 63-74.

Martin, C.K., Nacu, D., & Pinkard, N. (2016). Revealing Opportunities for 21st Century Learning: An Approach to Interpreting User Trace Log Data. *Journal of Learning Analytics,* 3(2): 37-87.

Mitchell, C. M., Boyer, K. E., & Lester, J. C. (2013, July). A Markov decision process model of tutorial intervention in task-oriented dialogue. In International Conference on Artificial Intelligence in Education (pp. 828-831). Springer Berlin Heidelberg..

MMS Education. (2012). Survey of K-12 educators on social networking, online communities, and web 2.0 tools. Retrieved from http://www.edweb.net/fimages/op/reports/

Nacu, D., Martin, C. K., Schutzenhofer, M., & Pinkard, N. (2016). Beyond traditional metrics: Using automated log coding to understand 21st century learning online. Proceedings of the 3rd ACM Conference on Learning@Scale, 197–200.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States.* Washington, DC: The National Academies Press.

Pea, R. (2013, February). Learning sciences and learning analytics: Time for a marriage. *EDUCAUSE Learning Initiative.* Denver, CO.

Pellegrino, J. W., & Hilton, M. L. (Eds.). (2013). *Education for life and work: Developing transferable knowledge and skills in the 21st century.* National Academies Press.

Reich, J., Murnane, R., & Willett, J. (2012). The state of wiki usage in US K-12 schools leveraging web 2.0 data warehouses to assess quality and equity in online learning environments. *Ed. Researcher,* 41(1), 7–15.

Rideout, V. (2015). *The common sense census: Media use by tweens and teens*. Common Sense Media. Retrieved from https://www. commonsensemedia.org/

Rodríguez, F. J., & Boyer, K. E. (2015, June). Discovering individual and collaborative problem-solving modes with hidden Markov models. In Conati, C., et al. (Eds.), *Artificial Intelligence in Education*. Paper presented at the 17th International Conference on Artificial Intelligence in Education (pp. 408-418). Springer International Publishing.

Siemens, G. (2012, April). Learning analytics: envisioning a research discipline and a domain of practice. In Shum, S., Gasevic, D., & Ferguson, R. (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* . Paper presented at the 2nd International Conference on Learning Analytics and Knowledge (pp. 4-8). ACM.

Taraghi, B., Saranti, A., Ebner, M., & Schön, M. (2014, June). Markov chain and classification of difficulty levels enhances the learning path in one-digit multiplication. In Zapharis, P., & Ioannou (Eds.), *Learning and Collaboration Technologies: Designing and Developing Novel Learning Experiences*. Paper presented at 1st Intl. Conference on Learning and Collaboration Technologies (pp. 322-333).

Warschauer, M., & Matuchniak, T. (2010). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of Research in Education*, 34(1), 179-225.

Watkins, C.S. (2011). *Digital divide: Navigating the digital edge.*

Xu, B. (2011). *Clustering educational digital library usage data: Comparisons of latent class analysis and K-Means algorithms.*

Zywica, J., Richards, K. A., & Gomez, K. (2011). Affordances of a scaffolded-social learning network. *On the Horizon*, 19(1), 33-42.

Acknowledgments

This material is based on work supported by the National Science Foundation Cyberlearning and Future of Learning Technologies grant (DIP #1325004).