REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
13-10-2017	Final Report	1-May-2014 - 30-Apr-2017

4. TITLE AND SUBTITLE

Final Report: Social Learning Networks: From Data Analytics to Active Sensing (Proposal for social informatics program, PM John Lavery)

5a. CONTRACT NUMBER
W911NF-14-1-0190

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHORS

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

Princeton University
PO Box 36
87 Prospect Avenue, Second Floor
Princeton, NJ 08544 -2020

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
65351-MA.6

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UU	UU	UU	UU		Mung Chiang
					19b. TELEPHONE NUMBER
					609-258-5071

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

RPPR Final Report
as of 13-Oct-2017

Agency Code:

Proposal Number: 65351MA **Agreement Number: W911NF-14-1-0190**

INVESTIGATOR(S):

Name: Mung Chiang
Email: chiangm@princeton.edu
Phone Number: 6092585071
Principal: Y

Organization: **Princeton University**
Address: PO Box 36, Princeton, NJ 085442020
Country: USA

DUNS Number: 002484665 EIN: 210634501
Report Date: 31-Jul-2017 Date Received: 13-Oct-2017
Final Report for Period Beginning 01-May-2014 and Ending 30-Apr-2017
Title: Social Learning Networks: From Data Analytics to Active Sensing (Proposal for social informatics program, PM John Lavery)
Begin Performance Period: 01-May-2014 **End Performance Period:** 30-Apr-2017
Report Term: 0-Other
Submitted By: Mung Chiang Email: chiangm@princeton.edu
Phone: (609) 258-5071
Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 1 **STEM Participants:** 1

Major Goals: A Social Learning Network (SLN) is a type of social network that exists between students, instructors, and modules of learning. Between 2015 and 2017, we completed a variety of data-driven research projects pertaining to SLN. Our work can be divided into three main thrusts: Predictive Learning Analytics (PLA), Social Learning Optimization (SLO), and Deep Learning Personalization (DLP).

The goal of our work on PLA was to develop algorithms to predict student outcomes. For SLO, we sought to study student-to-student and instructor-to-student interactions using the discussion forums from dozens of Massive Open Online Course (MOOC) offerings. In SLO, we have studied student-to-student and instructor-to-student interactions using the discussion forums from dozens of Massive Open Online Course (MOOC) offerings. Finally, building upon the methods developed in the PLA and SLO thrusts, the purpose of DLP was to develop systems for personalized learning that seek to maximize outcomes for each student.

Accomplishments: A Social Learning Network (SLN) is a type of social network that exists between students, instructors, and modules of learning. Between 2015 and 2017, we completed a variety of research projects pertaining to SLN, using data we obtained from several sources, including Massive Open Online Courses (MOOCs) that we taught, open source discussion forums we scraped, and corporate training courses we hosted. Our work on SLN can be divided into three main thrusts: Predictive Learning Analytics (PLA), Social Learning Optimization (SLO), and Deep Learning Personalization (DLP).

Our work on PLA has focused on developing algorithms to predict student outcomes in online courses. This has previously been challenging because intermediate performance data on learners (e.g., quiz or exam scores) tends to be sparse and only available at fixed points in time, making it difficult for algorithms to provide quality predictions in advance. In one of our projects, we developed two frameworks by which student video-watching clickstreams can be represented: one based on the sequence of events created, and another on the sequence of positions visited. With the event-based framework, we mined recurring subsequences of student behavior, which were found to contain fundamental characteristics such as reflecting (i.e., repeatedly playing and pausing) and revising (i.e., plays and skip backs). We call these behavioral motifs. Importantly, we found that some of these motifs were significantly correlated with changes in the likelihood that a student will be Correct on First Attempt (CFA) or not in answering quiz questions, and in ways that were not necessarily intuitive. Then, with the position-based framework, we devised models of quiz performance based on positions visited in a video. In evaluating these models through CFA prediction, we found that three of them could substantially improve prediction quality by up to 20%, underscoring the ability to relate this type of behavior to quiz scores.

Since these predictions considered videos individually, these benefits also suggested that behavioral data is useful in situations where there is limited training data, e.g., for early detection or in short courses. To this end, we also studied learning outcome prediction for online short-courses that have single outcomes assigned by instructors at the end. The lack of performance data makes the behavior of learners, captured as they interact with course content and with one another, essential for prediction. Our method defined several machine learning features based on the processing of behaviors collected on the modes of human learning in a course, including the previously identified motifs, and uses them in appropriate classifiers. Through evaluation on data captured from three two-week courses hosted through our delivery platforms, we make three key observations: (i) behavioral data contains signals predictive of learning outcomes in short-courses (with classifiers achieving AUCs above 80% after the two weeks), (ii) early detection is possible within the first week (AUCs exceeding 70% with the first week of data), and (iii) the content features have an “earliest” detection capability (with higher AUC in the first few days), while the social features become the more predictive set over time as the network matures.

In SLO, we studied student-to-student and instructor-to-student interactions using the discussion forums from dozens of MOOC offerings. One of the key challenges faced by these forums to-date is extremely low instructor-to-student ratios, motivating the development of methods to optimize participation. To do this, we first investigated the dynamics of user discussions, identifying course factors that are commonly associated with a decline in participation over time. Were able to verify that MOOC has a high decline rate in participation over time, and a high-volume of discussions that is infeasible for a teaching staff to manage itself. This motivated our more specific investigations into identifying factors associated with these decline rates, designing algorithms for thread relevance ranking, and modeling information propagation and efficiency in an SLN.

In particular, given the low teacher-to-student ratios in MOOC, the underlying efficacy of the SLN in these scenarios hinges on the notion that strong social ties will form between those seeking information on certain course topics and those who are providing information on these same topics. To study this, we developed methodology for quantifying the efficiency of information propagation in discussion forums. Under our method, each student is modeled individually as possessing certain levels of seeking (i.e., question asking) and disseminating (i.e., question answering) tendencies on a set of (latent) topics that define the discussions in the course, and socially based on the probability that she responds when another particular student makes a post. Efficiency is quantified by comparing the amount of utility present in the observed SLN to a benchmark of maximum utility achievable through network optimization subject to a set of (inferable) constraints.

This SLN optimization posed two computational challenges that we needed to overcome in our implementation. First is one of scalability: it has several million variables corresponding to the weights in a directed user-to-user graph, making it intractable for standard convex optimization solvers. As a result, we derived a projected gradient descent algorithm for this problem in which the projection step becomes quadratic, and in turn derive a proximal alternating direction method of multipliers (ADMM) algorithm to perform the projection. The second challenge is non-uniqueness: the optimization problem has non-unique solutions for realistic parameter values, posing the question of how to obtain the most practical solution. We evaluated the efficiency of several SLN based on our method. In doing so, we made a number of interesting findings, such that (i) SLN efficiency in MOOC discussions can be rather low (from 76% to 90% depending on the specific parameters and dataset), indicating that much can be gained through optimization, (ii) the gains in global utility can be obtained without making the distribution of local utilities less fair, and (iii) the optimized SLN introduces a plethora of connections between users that were not present beforehand.

Building upon the methods developed in the PLA and SLO thrusts, in DLP we have developed systems for personalized learning that seek to maximize outcomes for each student. The systems collect behavioral data about students as they interact with course content and with one another, and use that data for real-time updating of user models that in turn dictate the learning path of each student. In particular, we have designed, implemented, and evaluated our Adaptive Educational System (AES) called the Mobile Integrated and Individualized Course (MIIC). MIIC is a platform for personalized course delivery which integrates lecture videos, text, assessments, and social learning into a mobile native app, and collects clickstream-level behavioral measurements about each student as they interact with the material. Recruiting students from one of our MOOCs, we have conducted two preliminary trials with MIIC, in which we found (i) that the majority of students (70 percent) preferred MIIC overall to a one-size-fits-all (OSFA) presentation of the same material, (ii) that the mean level of engagement, when quantified as the number of pages viewed, was statistically higher (by 72 percent) among students using MIIC than among OSFA, and (iii) that the integrated multimedia learning features were generally favorable among the students (e.g., 87 percent found the videos helpful).

Training Opportunities: Nothing to Report

Results Dissemination: Journal paper publications:

C. Brinton, S. Buccapatnam, M. Chiang, H. V. Poor. Mining MOOC Clickstreams: Video-Watching Behavior versus In-Video Quiz Performance.
IEEE Transactions on Signal Processing, Vol. 64, No. 14, p. 3677-3692, 2016.

C. Brinton, R. Rill, S. Ha, M. Chiang, R. Smith, W. Ju. Individualization for Education at Scale: MIIC Design and Preliminary Evaluation. IEEE Transactions on Learning Technologies, Vol. 8, No. 1, p. 136-148, 2015.

C. Brinton, M. Chiang, S. Jain, H. Lam, Z. Liu, F. Wong. Learning about social learning in MOOCs: From statistical analysis to generative model. IEEE Transactions on Learning Technologies, Vol. 7, No. 4, p. 346-359, 2014.

Conference paper publications:

W. Chen, C. Brinton, D. Cao, M. Chiang. Behavior in Social Learning Networks: Early Detection for Online Short-Courses. IEEE INFOCOM, 2017.

C. Brinton, S. Buccapatnam, F. Wong, M. Chiang, H. V. Poor. Social Learning Networks: Efficiency Optimiza- tion in MOOC Forums. IEEE INFOCOM, 2016.

C. Brinton, M. Chiang. MOOC Performance Prediction via Clickstream Data and Social Learning Networks. IEEE International Conference on Computer Communications (INFOCOM), 2015.

C. Brinton, M. Chiang. Social Learning Networks: A Brief Survey. 48 Annual Conference on Information Science and Systems (CISS), 2014.

Honors and Awards: Christopher G. Brinton received the Bede Liu Best Dissertation Award in Electrical Engineering in May 2016 for his PhD thesis on Social Learning Networks.

Christopher G. Brinton received the Yan Huo ’94 Graduate Fellowship in Electrical Engineering in December 2015.

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: Graduate Student (research assistant)
Participant: Christopher Greg Brinton
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Swapna Buccapatnam
Person Months Worked: 10.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N

Other Collaborators:

ARTICLES:

Publication Type: Journal Article Peer Reviewed: **Publication Status:** 1-Published
Journal: IEEE Transactions on Knowledge and Data Engineering
Publication Identifier Type: Publication Identifier:
Volume: 28 Issue: 8 First Page #: 2158
Date Submitted: Date Published: 8/2/16 2:25PM
Publication Location:
Article Title: Quantifying political leaning from tweets, retweets and retweeters
Authors: F. M. F. Wong, S. Sen, C. W. Tan, and M. Chiang
Keywords: Twitter, political science, data analytics, inference, convex programming, signal processing
Abstract: The widespread use of online social networks (OSNs) to disseminate information and exchange opinions, by the general public, news media and political actors alike, has enabled new avenues of research in computational political science. In this paper, we study the problem of quantifying and inferring the political leaning of Twitter users. We formulate political leaning inference as a convex optimization problem that incorporates two ideas: (a) users are consistent in their actions of tweeting and retweeting about political issues, and (b) similar users tend to be retweeted by similar audience. We then apply our inference technique to 119 million election-related tweets collected in seven months during the 2012 U.S. presidential election campaign. On a set of frequently retweeted sources, our technique achieves 94% accuracy and high rank correlation as compared with manually created labels. By studying the political leaning of 1,000 frequently retweeted sources, 232,000 etc
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

BOOKS:

Publication Type: Book Peer Reviewed: Y **Publication Status:** 1-Published
Publication Identifier Type: ISBN Publication Identifier:
Book Edition: 1st Volume: Publication Year: 2016 Date Received: 02-Nov-2016
Publication Location:
Publisher: Princeton University Press
Book Title: The Power of Networks
Authors: C. Brinton, M. Chiang
Editor:
Acknowledged Federal Support: Y

CONFERENCE PAPERS:

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: IEEE INFOCOM
Date Received: 02-Nov-2016 Conference Date: 10-Apr-2016 Date Published:
Conference Location: San Francisco
Paper Title: Social learning networks: Efficiency optimization for MOOC forums
Authors: C. Brinton, S. Buccapatnam, M. Chiang, H. V. Poor
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 13-Oct-2017

Nothing to report in the uploaded pdf (see accomplishments)