Pass the Idea Please: The Relationship between Network Position, Direct Engagement, and Course Performance in MOOCs

Stacey L. Houston, II
Vanderbilt University
Department of Sociology
stacey.houston@vanderbilt.edu

Katherine Brady
Vanderbilt University
Department of Electrical Engineering and Computer Science
katherine.a.brady@vanderbilt.edu

Gayathri Narasimham
Vanderbilt University
Vanderbilt Institute for Digital Learning
gayathri.narasimham@vanderbilt.edu

Douglas Fisher
Vanderbilt University
Department of Electrical Engineering and Computer Science
douglas.h.fisher@vanderbilt.edu

ABSTRACT

Extant research suggests that learner engagement in discussion forums is positively correlated with learner performance. In this paper we investigate which types of forum engagement are most strongly associated with final performance in MOOC courses. In particular, we compare the correlation between course final grade and two types of learner engagement: direct measures, which count the number of interactions, and indirect measures, which capture learners position in a social network. We found that direct measures have stronger correlations with final grade. However, in preliminary analyses, we also found that course instructors score higher than learners on some indirect measures. We discuss the implications of these findings and our plans for developing the work further in the future.

Author Keywords

MOOCs; Social Network Analysis; Course Performance

ACM Classification Keywords

K.3.1 Computer Uses in Education: Distance learning

INTRODUCTION

Models which treat the spread of information as a type of contagion have been verified by historical records [1, 7]. This, along with studies that have found word of mouth more effective than other forms of marketing [15, 11], provide strong evidence that interacting with classmates should increase learning in a course. In this paper we investigate whether learners benefit more from interacting directly with peers or positioning themselves to indirectly receive knowledge and ideas through the forum's social network.

We investigate whether learners' final grades are more closely correlated with direct or indirect connections to their peers. To capture these connections, we modeled learner interactions on course discussion forums for three massive open online courses (MOOCs) on the Coursera Platform as a social network. The **direct measures** were the number of threads on which a learner posted and the number of peers a learner interacted with on those threads. The **indirect measures** we explored were the frequency at which a learner serves as the link between peers (betweenness), a learner's connectedness to influential peers (Bonacich Power), and the number of peers through which a message from a learner would need to travel to reach all other learners (closeness).

Our findings indicate that direct measures of learner interactions are more consistently correlated with final grade than a learner's indirect interactions. However, we found some evidence that indirect metrics may be better suited for identifying the most influential members of a forum. Expanding on this evidence will be necessary for drawing any concrete conclusions.

BACKGROUND AND RELATED WORK

Researcher interest in the ways social networks affect learning has been around for some time. Martinez et al. established in 2003 that networks of social interactions in traditional classrooms are a good model of learner collaboration [10]. As online classrooms increasingly developed active forums which automatically record student interactions, methods were tailored for building social networks from forum data [9]. Research has demonstrated that, as in traditional classrooms, peer interactions in online discussion forums can foster learning [6]. Furthermore, more active networks pro-

L@S 2017, April 20–21, 2017, Cambridge, MA, USA
 ISBN 978-1-4503-4450-0/17/04...$15.00
DOI: http://dx.doi.org/10.1145/3051457.3054008

mote higher levels of cognitive engagement and critical thinking [14].

There has been some research into the type of interactions that best foster learning. A study that investigated networks of scientific publications suggests that networks with large diameters and many small tightly connected communities (like a company with many small departments) spread ideas best [4]. Studies which have looked at the benefits to individual learners have concluded, like us, that direct metrics like the number of posts a learner makes or the number of peers they interact with, are the most predictive of performance metrics such as dropping out or completing [16, 13]. These studies explored whether social statistics can predict learner outcomes. Our work differs from these studies in asking whether *some* statistics are better predictors than others.

Though our work suggests that indirect links are not as predictive of performance outcomes, research suggests closeness is correlated with having more influence within the network. Cho et al. found through surveys of an online master's program that learners who were close on the friendship network before the beginning of a course had a larger influence on how the social network developed during the course [2]. Dowell et al. found that learners in a MOOC who were close used longer, more narrative form language [3]. This findings suggests that measures of indirect interaction may have implications beyond course performance. The preliminary component of this study regarding course instructors intends to develop this line of thinking.

The present study contributes to the existing literature in three ways:

1. It utilizes data from three different courses to replicate findings from existing literature. In addition, we examined different time points in each course to determine the extent to which the timing of course interactions might influence, or moderate, the relationship between learner interactions and final course performance.

2. It compares the association between course performance and two specific types of learner interaction on online discussion forums: direct and indirect. Previous studies have tended to examine these types separately, neglecting to compare the usefulness of exploring one type of interaction over another.

3. It offers a preliminary look at instructor positioning in the social network. Given the implicit assumption that course instructors are most expert among the course participants, exploring the ways in their position differs from learners' is a fruitful avenue for continued research.

DATASET

The data used in this research came from three MOOC sessions. Two of these sessions were different offerings of Innovation, a course on entrepreneurship which heavily encouraged forum participation. The other session was Matlab, a course on programming. All three sessions had active instructor and teaching assistant involvement on the forums.

Course	Sign Ups	Assignment	Unique Posters
Innovation 1	47,000	7,200	6,100
Innovation 2	33,000	3,300	2,000
Matlab 1	63,000	7,700	1,500

Table 1: Number of learners in each course who signed up, did at least one assignment or posted on the forum.

Course	Sub-Forum	Posters	Threads
Innovation 1	Week 1	4329	532
Innovation 1	Week 4	1798	247
Innovation 1	Week 8	1236	112
Innovation 2	Week 1	1493	327
Innovation 2	Week 4	811	280
Innovation 2	Week 8	435	77
Matlab 1	Weeks 1-4	700	378
Matlab 1	Week 5	163	79
Matlab 1	Week 8	266	131

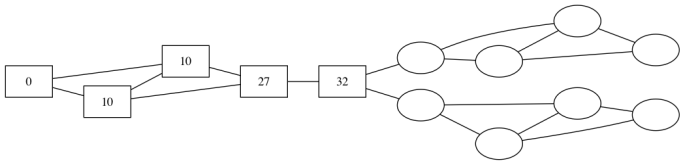
Table 2: This table shows the number of posters (including both learners and instructors) who participated in each subforum and the number of threads they generated in that subforum.

Course announcements were made on a separate announcements page, not the forums.

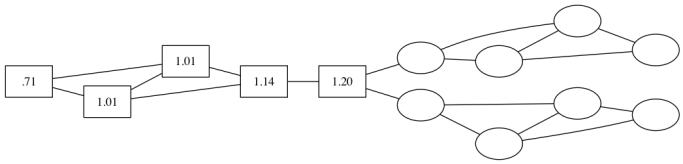
It is important to consider time in any analysis of the relationship between course interaction and grade. Learners who post in a discussion forum about the final project in a course are more likely to receive passing grades than students who post about the first lecture since some of the learners who watch the first lecture drop out before the final project [6]. The MOOC sessions under study here all had separate sub-forums for each week of the course. We take this into consideration in this study. Specifically, in each course, we modeled subforums from the beginning, middle and end as separate social networks rather than treating the whole forum as single social network. The number of participants in each session and sub-forum are shown in tables 1 and 2.

Method

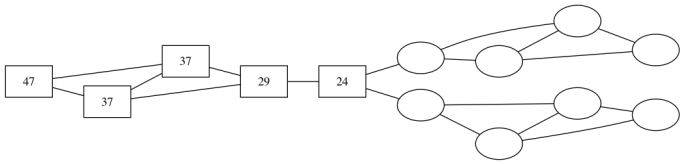
In these MOOCs, after a learner made a post on a thread they were automatically added to a mailing list which would send out an email each time there was a subsequent new post on that thread. For this reason we modeled learners as being exposed to every post on threads to which they posted. To represent this as a social network, we modeled every learner as a node. Two nodes shared an edge if the corresponding learners had co-participated in at least one thread in a subforum. This method of modeling the forums has been corroborated by other researchers [8, 12, 6]. A popular alternative is to model the discussion forums as directed graphs [3, 9, 16] where edges point to the peer a learner is replying to. For the forums explored in this study there was no reliable means of recording if one post was in reply to another. Therefore, we used a model with undirected edges. To account for the difference in course commitment between learners who have



(a) The value inside each node shows the betweenness of that node.



(b) The value inside each node shows the Bonacich Power of that node.



(c) The value inside each node shows the closeness value of that node. A node is considered "closer" if it has a lower value.

Figure 1: The above graph shows examples of the values for each indirect measure we looked at on a small example network.

just started the course and learners in the last week, we did not combine models across sub-forums.

The purpose of this study is to determine the relative correlation between grade and metrics of direct connections versus indirect metrics. We differentiate between direct learner connections and indirect learner connections based on the manner in which information is spread to learners. Direct metrics capture the extent to which one learner is exposed to the ideas or knowledge of another learner. Indirect metrics, on the other-hand, capture the extent to which a learner positions themselves to be exposed to a variety of other learners. Further, direct metrics are captured by counts of user activity, whereas indirect metrics must be calculated from social network models. The specific direct metrics we examined were the number of threads to which a learner posted and the number of peers with whom they shared at least one thread. We contrasted these with three indirect metrics of social network positioning: Betweenness, Bonacich Power and Closeness. Betweenness measures how often a learner lies on the shortest path between two peers. Bonacich Power is a reciprocal measure in which the connectedness of a learner's neighbors on the network determines their Bonacich Power. Closeness measures the sum of the shortest paths from a learner to each of their peers - thus, learners with lower closeness scores are more central. The values of all three of these metrics are shown in figure 1.

To provide examples of the indirect measures, a learner in a MOOC course with high betweenness is one who frequently engages with posts by peers from different groups, thereby hypothetically exposing themselves to a variety of diverse ideas. A learner with high Bonachich Power posts in threads

Course	Sub-Forum	Neigh	Thread
Innovation 1	Week 1	ns	0.08**
Innovation 1	Week 4	0.07**	0.05*
Innovation 1	Week 8	0.06*	ns
Innovation 2	Week 1	0.10**	0.16**
Innovation 2	Week 4	ns	0.08*
Innovation 2	Week 8	ns	ns
Matlab 1	Weeks 1-4	0.16**	0.15**
Matlab 1	Week 5	0.20*	0.15*
Matlab 1	Week 8	ns	ns

Table 3: The significant correlations with grade for direct interactions: number of neighboring peers (Neigh) and number of threads posted to (Thread). In this table ns = not significant, * = $p < .05$ and ** = $p < .01$

Course	Sub-Forum	Close	Bon	Bet
Innovation 1	Week 1	-0.05**	ns	0.05**
Innovation 1	Week 4	-0.09**	ns	ns
Innovation 1	Week 8	-0.09**	ns	ns
Innovation 2	Week 1	-0.12**	ns	0.09**
Innovation 2	Week 4	ns	ns	0.07*
Innovation 2	Week 8	ns	ns	ns
Matlab 1	Weeks 1-4	ns	ns	0.14**
Matlab 1	Week 5	ns	0.16*	ns
Matlab 1	Week 8	ns	ns	ns

Table 4: The significant correlations with grade for indirect interactions: Betweenness (Bet), Closeness (Close) and Bonacich Power (Bon). In this table ns = not significant, * = $p < .05$ and ** = $p < .01$

that are frequented by other highly engaged and active learners, thereby hypothetically exposing themselves to peers who have learned from others in the course. Lastly, a learner with a low closeness score interacts on threads in a manner that exposes them to ideas of many other learners in the course. The results of our analyses comparing these metrics are presented in the following section.

RESULTS

The number of threads a learner posted to was significantly correlated with grade more often than any of the other statistics (in 6 out of the 9 sub-forums). Furthermore, those six sub-forums are the ones from the beginning and middle of each course, suggesting that this statistic is most useful at earlier points in the course.

The next most promising statistic was the number of neighboring peers a learner has followed by betweenness and closeness. Bonacich Power was only significant in one subforum. These results suggest that direct measures of learner interactions are more consistently correlated with course performance than indirect measures. The significant correlations for each measure can be seen in tables 3 and 4.

In analyses not presented here, we estimated a few models which used linear combinations of indirect statistics to see if

grade could be better predicted using multiple linear regression. None of these models explained more variation in learners final grade than number of threads. Thus, we concluded that combining the features does not add value if the outcome of interest is final course performance.

In addition to correlating each statistic with grades, we also explored the difference between the measurements for instructors and students. In the Matlab course, instructors were more engaged than the average learner, as measured by all the metrics we considered in this study. This was not true for the Innovation courses with the exception of closeness. In all sub-forums in which instructors participated, the instructor closeness scores were at least a standard deviation 'closer' than the mean closeness score. Dowel et al. found that forum participants who had high closeness scores tended to use more narrative language, much like instructors [3]. We hope to explore the implications of this in future work.

CONCLUSIONS AND FUTURE WORK

In this study we found that direct learner interactions on the forums are more often correlated with learners' final grades than indirect interactions. We hypothesize that this trend arises due to similar ideas being posted on multiple threads and plan to investigate the veracity of this hypothesis in future work. We also find that, in general, instructors are more indirectly connected than learners in the social networks.

In future work, we hope to further uncover ways in which the behaviors of learners are related to behaviors of instructors. Additionally we hope to do a qualitative assessment of the diversity of thread topics on each sub-forum. We would expect indirect measures to be more correlated with grade on sub-forums which rarely have the same ideas repeated on multiple threads and less correlated with grade on sub-forums where multiple threads contain similar ideas. Our current model treats all peers in the social network as more or less the same. However, research suggests that learners benefit more from interacting with peers with high GPAs [5]. We plan to look into whether interactions with instructors and teaching assistants lead to different benefits than peer interactions.

REFERENCES

1. Bettencourt, L. M., Cintrón-Arias, A., Kaiser, D. I., and Castillo-Chávez, C. The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Physica A: Statistical Mechanics and its Applications 364* (2006), 513–536.
2. Cho, H., Gay, G., Davidson, B., and Ingraffea, A. Social networks, communication styles, and learning performance in a cscl community. *Computers & Education 49*, 2 (2007), 309–329.
3. Dowell, N. M., Skrypnyk, O., Joksimovic, S., Graesser, A. C., Dawson, S., Gaševic, D., Hennis, T. A., de Vries, P., and Kovanovic, V. Modeling learners' social centrality and performance through language and discourse. *International Educational Data Mining Society* (2015).
4. Gasevic, D., Kovanovic, V., Joksimovic, S., and Siemens, G. Where is research on massive open online courses headed? a data analysis of the mooc research initiative. *The International Review of Research in Open and Distributed Learning 15*, 5 (2014).
5. Gašević, D., Zouaq, A., and Janzen, R. Choose your classmates, your gpa is at stake!: The association of cross-class social ties and academic performance. *American Behavioral Scientist* (2013), 0002764213479362.
6. Gillani, N., and Eynon, R. Communication patterns in massively open online courses. *The Internet and Higher Education 23* (2014), 18–26.
7. Gillani, N., Yasseri, T., Eynon, R., and Hjorth, I. Structural limitations of learning in a crowd: communication vulnerability and information diffusion in MOOCs. *Scientific reports 4* (2014), 6447.
8. Jiang, S., Fitzhugh, S. M., and Warschauer, M. Social positioning and performance in moocs. In *Workshop on Graph-Based Educational Data Mining* (2014), 14.
9. Laat, M., Lally, V., Lipponen, L., and Simons, R.-J. Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *International Journal of Computer-Supported Collaborative Learning 2*, 1 (2007), 87–103.
10. Martınez, A., Dimitriadis, Y., Rubia, B., Gómez, E., and De La Fuente, P. Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education 41*, 4 (2003), 353–368.
11. Nekovee, M., Moreno, Y., Bianconi, G., and Marsili, M. Theory of rumour spreading in complex social networks. *Physica A: Statistical Mechanics and its Applications 374*, 1 (2007), 457–470.
12. Poquet, O., and Dawson, S. Analysis of MOOC Forum Participation. In *Australasian Society for Computers in Learning and Tertiary Education* (2015), 224–234.
13. Romero, C., López, M.-I., Luna, J.-M., and Ventura, S. Predicting students' final performance from participation in on-line discussion forums. *Computers & Education 68* (2013), 458–472.
14. Thomas, M. J. Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning 18*, 3 (2002), 351–366.
15. Trusov, M., Bucklin, R. E., and Pauwels, K. Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site. *Journal of marketing 73*, 5 (2009), 90–102.
16. Yang, D., Sinha, T., Adamson, D., and Rosé, C. P. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-driven education workshop*, vol. 11 (2013), 14.