

Patterns of Interaction Predict Social Role in Online Learning

Huaiyi (Cici) Sun, Dan Calacci, and Alex (Sandy) Pentland

Massachusetts Institute of Technology
77 Massachusetts Ave, Cambridge, MA 02139, USA
sandy@media.mit.edu

Abstract. We identify key factors that influence the social role of students in an online class and predict student grades and roles using communication features. Our results show that students that communicate with others less overall are more likely to work alone on future projects, while future group leaders are more likely to have high rates of communication with other study groups. Interestingly, social patterns seem to be as important as learning performance and more important than demographic information in predicting a student's social role, and integral in predicting a student's grade.

Keywords: Online education, Second network, Social roles, Leader, Loner, Communication.

1 Introduction

Online learning environments increasingly include peer interaction as a core part of the learner experience [1-3]. As more learners engage with their peers using synchronous communication tools such as the text chat system studied here, it is important to understand how peer interactivity affects both the social roles of learners as well as their learning outcomes. Understanding how social roles affect distributed group work within small teams is of particular importance. We use chat data obtained from a course-specific chat platform over the course of a twelve-week online class of over 900 students to understand how student social roles relate to communication patterns and course performance.

We examine objective data on student's emergent social roles -- whether later in the class they will lead a group project (a "leader") or decide to work alone (a "loner") -- to better understand learner experience and outcomes. We find significant differences in how "leaders" and "loners" approach group interaction and social presence, and we find evidence suggesting that learning performance is related to peer engagement patterns. Specifically, we find that communication patterns in the first eight weeks of the course can be used to fairly accurately predict whether a student will lead a group or work alone.

2 Related Work

2.1 Loners and loneliness in social networks

Few prior studies have clearly articulated “loner” as a social role or identified key factors for predicting the existence of “loners” in a class or even in a general social network. However, loneliness as an emotional and social cognitive term [4] has been well documented in two types of research: one in child development, which mostly relates to the study of autism [5], and the other in social psychology. In these cases, the primary method of identifying loneliness is via self-reporting and questionnaires. For example, one study looked at the rate of loneliness as a self-reported feeling among high-functioning children with autism compared to children without autism [5]. Other studies also show that loneliness can be predicted by the quality of one’s social network as well as measures of personal characteristics such as extraversion and neuroticism [6]. Furthermore, how loneliness occurs and spreads in a social network has also been clearly deciphered [7].

2.2 Leaders and leadership in social networks

Few studies have been directed to identifying factors in predicting whether a student will be a leader in a classroom social network, however, outside the realm of classroom, research about leaders in a general social network shows that leaders tend to be more mature, especially in organizational studies [8]. Previous studies have shown that a leader’s mood plays an important role in group performance [9]; Social network ties such as the centrality in communication network is correlated with the emergence of leadership and group effectiveness [10-11]; Leader-member exchange theory (LMX) pays attention to the dyadic link between a leader and a subordinate with attention to how this relationship affects the subordinate’s performance [12].

Therefore, our study of student social roles in an online class network breaks new ground and connects to previous research on both loneliness and leadership in a general social network.

3 Data

The data we used were collected in 2016 from a twelve-week online mid-career finance class with over 900 students at the MIT Sloan School of Management. The raw data can be divided into three types: group activity data, online text chat data, and demographic data. (1) *Group activity data*: The first eight weeks of the course consisted of weekly modules and group assignments. Students were randomly assigned to “Exercise” study groups at the beginning of the course, and they received individual grades for each weekly assignment. In the last 4 weeks of the course, students self-organized into “Capstone” project groups after viewing project proposals from other students in the class. In these last 4 weeks, each group received a weekly grade based on their group project performance. All grade data were on a 0-100 scale. Our data consists of a total

of 192 “Exercise” study groups, each consisting of 5-6 group members, and 252 “Capstone” groups. As Capstone groups were self-organized, they vary widely in size. Some groups consist of only one student; the largest group in our dataset has 24 students. (2) *Online Communication Data*: Throughout the course period, students could send synchronous text chat messages via a course-specific online chat platform, similar to Slack. Students were able to join pre-existing channels created by course staff, send direct messages to other users, and create their own channels. Students sent a total of over 50,000 messages in over 800 channels. For each message sent in the platform, our dataset includes a unique identifier for the channel and sender, the message content, and the timestamp of the message. (3) *Demographic Data*: We also have self-reported demographic data from students, including age ($M=39.81$, $SD=8.80$), length of work experience, geographic region, level of education, professional industry, and language.

4 Method

Loners and Leaders: We first characterized students as either “Loners” or “Leaders”. Loners are students that chose to work alone on a Capstone class project. Leaders are students that proposed an original capstone project idea that their group decided to work on, and who coordinated their group’s submissions during the last four weeks of the class. We identify a total of 86 loners and 135 leaders.

Features Extraction: We computed a variety of communication, performance, and demographic features for each student using data from the first eight weeks of the class (the period during which students were randomly assigned to Exercise project groups). These features would lead up to Capstone project group formation, part of them are for example: (1) *Total msg*: total message counts of a student. (2) *Intra msg*: message counts of a student sent only to Exercise group members. (3) *Inter msg*: message counts of a student sent only to non-Exercise group members. (4) *Known ct*: number of unique students’ user sent a message to. (5) *Grade*: average grade by week. (6) *English*: native language of the student. **Table 1** shows the summary statistics for these features.

Table 1. Subset of Feature Statistics from first eight weeks of course

	Grade	Age	Total msg	Intra msg	Inter msg	Known ct
Min	0	17	1	0	0	1
Max	100	70	3366	2036	1330	89
Mean	85.71	39.83	53.24	37.50	15.74	11.30

Modeling: To test our hypotheses, we used a simple logistic regression model to identify which variables are related to a student’s probability of being a loner or leader. After visually validating our modeling results, we developed a Random Forest classification model to investigate the predictive power of our computed features. Our dataset

is highly imbalanced, so we used Synthetic Minority Over-sampling (SMOTE) to improve predictive accuracy between our two classes. We also used a similar approach to investigate the significance of communication patterns in predicting student grades.

5 Results

Results from the logistic regression in **Table 2** suggest that students who achieve lower grades, communicate with fewer people, and are less active in chat are more likely to work alone in the Capstone period of the class. Most demographic features such as age, language, or geographic region do not seem to carry a significant effect. The only significant (at the $p < 0.05$ level) demographic effect we observe is that students with a self-reported geographic region of “Asia” are more likely to work alone. To conclude, students with less active communication patterns are more likely to become “loners”.

Table 2. Loner.

	Estimate	SE	Z value	p-value	[95% CI]	
Grade	-0.0306	0.005	-6.517 ***	0.000	-0.040	-0.021
Age	-0.0068	0.007	-0.949	0.343	-0.021	0.007
Asia	-0.4299	0.129	-3.323 ***	0.001	-0.683	-0.176
Education	-0.4627	0.269	-1.722	0.085	-0.989	0.064
English	-0.0501	0.141	-0.355	0.722	-0.326	0.226
Total message	-0.5733	0.048	-11.84 ***	0.000	-0.668	-0.478
Known count	-0.0155	0.006	-2.519 *	0.012	-0.028	-0.003

* $p < .05$, ** $p < .01$, *** $p < .001$

Leaders, on the other hand, are more likely to have more active communication patterns. The regression results in **Table 3** suggest that students who send more inter-group messages are more likely to become group leaders, while frequency of within-group communication has no effect. Interestingly, communicating with other groups (higher inter-group message count) seems to be as significant as a students’ grade in predicting whether they will lead a project group.

Table 3. Leader

	Estimate	SE	Z value	p-value	[95% CI]	
Grade	0.0943	0.014	6.959 ***	0.000	0.068	0.121
Age	-0.0105	0.007	-1.499	0.134	-0.024	0.003
Asia	-0.3327	0.138	-2.413*	0.016	-0.603	-0.062
Education	0.2846	0.273	1.043	0.297	-0.025	0.820
English	0.6001	0.127	4.715***	0.000	-0.326	0.226
Inter-group message	0.3927	0.057	6.841***	0.000	0.280	0.505
Intra-group message	-0.0637	0.047	-1.354	0.176	-0.156	0.028

*p < .05, ** p<.01, *** p<.001

Figure 1 and **Figure 2** show the probability density of the grade and message counts for both loners and non-loners. Loners are much more likely to have very low message counts in comparison with other students, and the distribution of student grades for loners is much more dispersed than for non-loners. **Figure 3** and **Figure 4** demonstrate the model performance of the random forest in predicting loners and leaders and obtaining encouraging results: AUC of 0.75 and AUC of 0.73. **Table 4** shows the results of the grade prediction model with an *R square value of 0.24*. The top variables included in **Table 4**, indicate that English speakers who have strong communication within their course group are more likely to receive high grades. This result considered together with our leader model indicates that while effective inter-group communication is associated with leadership, within-group communication is more associated with high achievement. This makes sense: student's grades are highly dependent on their group performance, while to lead a group, students must rally others from the class.

Table 4. Grade Prediction

	Estimate	SE	Z value	p-value	[95% CI]	
Known Count	-0.0291	0.086	-0.338	0.736	-0.199	0.141
Age	-0.2340	0.125	-1.879	0.062	-0.480	0.012
Asia	-7.0331	2.591	-2.715**	0.007	-12.144	-1.922
Education	-2.9315	4.430	-0.662	0.509	-11.670	5.807
English	10.8115	2.431	4.448***	0.000	6.017	15.606
Inter-group message	0.2581	0.910	0.284	0.777	-1.538	2.054
Intra-group message	2.1833	0.785	2.781**	0.006	0.635	3.732

*p < .05, ** p<.01, *** p<.001

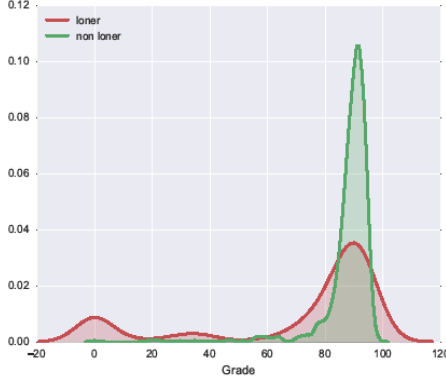


Fig. 1. PDF for Grade

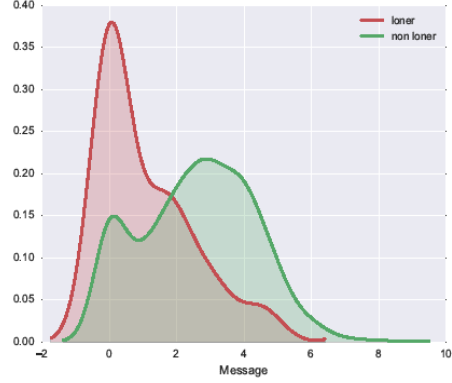


Fig. 2. PDF for Message

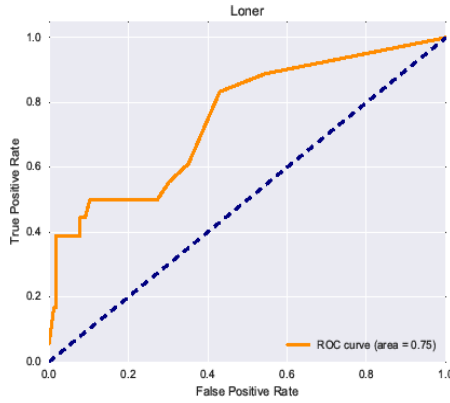


Fig. 3. AUC curve for Loner prediction

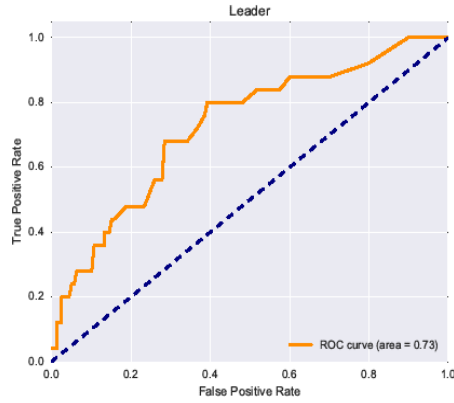


Fig. 4. AUC curve for Leader prediction

6 Discussion & Conclusion

Perhaps unsurprisingly, we find that project leaders exhibit a higher degree of social presence than non-leaders or loners. However, higher overall social engagement -- i.e., sending more overall messages -- is not strongly associated with leader behavior. Instead, project leaders tend to discuss and interact with many students from *different groups*, mirroring our earlier result in face-to-face college class projects [13] and suggesting that students who discuss with a higher diversity of people tend to both be higher-performing and more likely to lead teams.

Loners, students who do an independent group project, tend instead to be more socially isolated throughout the course and achieve overall lower grades than leaders or non-loners. Interestingly, our analysis suggests that the most significant variable in determining whether a student works alone is their overall social engagement with others

in the class -- the normalized number of messages they send. Users who are more socially engaged, even if only with their group members, are less likely to work alone.

These results suggest that in classes that focus on group learning, peer engagement outside of a student's assigned work group is an important predictor of success and achievement. Our work suggests that by using data on student interaction, educators might be able to intervene mid-course to help isolated students retain more knowledge and engage with course material more effectively. Simple metrics, like the number of messages a student has sent to others, or how diverse their communication is, can be used to craft curriculum that serves students more effectively.

References

1. Calacci, D, et al: Breakout: An open measurement and intervention tool for distributed peer learning groups. arXiv preprint arXiv:1607.01443 (2016).
2. Pentland,A: The new science of building great teams. *Harvard Business Review*,90.4 (2012)
3. Poole, D.M: Student participation in a discussion-oriented online course: A case study. *Journal of research on computing in education* 33.2. 162-177 (2000).
4. Weiss, S. Loneliness: The experience of emotional and social isolation. (1973).
5. Bauminger,N,etal: Loneliness and friendship in high-functioning children with autism. *Child development* 71.2 .447-456 (2000).
6. Stokes,P: The relation of social network and individual difference variables to loneliness. *Journal of Personality and Social Psychology* 48.4 .981 (1985).
7. Nicholas.C. Alone in the crowd: The structure and spread of loneliness in a large social network. *Journal of personality and social psychology* 97.6. 977 (2009).
8. Mehra, A,et al:The social network ties of group leaders: Implications for group performance and leader reputation. *Organization science* 17.1. 64-79 (2006).
9. Sy, T: The contagious leader: impact of the leader's mood on the mood of group members, group affective tone, and group processes. *Journal of applied psychology* 90.2. 295 (2005).
10. Bavelas, A: Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*22.6. 725-730 (1950).
11. Leavitt, J: Some effects of certain communication patterns on group performance. *The Journal of Abnormal and Social Psychology* 46.1. 38 (1951).
12. Graen, G: A role-making model of leadership in formal organizations: A developmental approach. *Leadership frontiers* 143. 165 (1975).
13. Yves-Alexandre, et al: The strength of the strongest ties in collaborative problem solving. *Scientific reports* 4. 5277 (2014).