Modeling Network Dynamics of MOOC Discussion Interactions at Scale

Jingjing Zhang
Beijing Normal University
Beijing, China
jingjing.zhang@bnu.edu.cn

Maxim Skryabin
Stepik
Saint Petersburg, Russia
ms@stepik.org

ABSTRACT
This paper attempts to model network dynamics of MOOC discussion interactions. It contributes to providing alternatives to conducting null hypothesis significance testing in educational studies. Using data collected from two successive psychology MOOCs in 2014 and 2015, the probabilistic longitudinal network analysis was performed by employing stochastic actor-based models with statistical accuracy. Understanding the mechanisms that drive the dynamics of discussions shed light on the design of a self-generated and learner-supported learning environment to meet the challenges of accommodating a massive and global student body.

1. Author Keywords

interactions; SIENA; probabilistic longitudinal network analysis; network dynamics; peer-supported learning.

2. ACM Classification Keywords

I.6.4 Simulation and Modelling: Model Validation and Analysis.

INTRODUCTION

Understanding learning at scale is a challenging task. As stated earlier, particular concerns are the extremely high rates of attrition and the pattern of steeply unequal participation in MOOCs. Using traditional educational methods fail to link the observed behavioral patterns within a network to the underlying the effects of network structure and the role of the participants that may explain why these patterns emerge. This study is an empirical investigation of the network dynamics of MOOC discussions, and attempts to make a contribution to providing alternatives to conducting null hypothesis significance testing in educational studies. Understanding the mechanisms that drive the dynamics of discussions shed light on the design of a self-generated and learner-supported learning environment to meet the challenges of accommodating a massive and global student body.

Using data collected from two successive psychology MOOCs in 2014 and 2015 and applying probabilistic longitudinal network analysis, this study seeks to rigorously measure the dynamic mechanisms that drive discussion change over time. The probabilistic analysis was performed by employing stochastic actor-based models with statistical accuracy.

METHODS

The probabilistic longitudinal network analysis was performed by employing stochastic actor-based models defined and evaluated with the program Simulation Investigation for Empirical Network Analysis. Four hypotheses are proposed to test the network dynamics of MOOC discussions.

Hypothesis 1 (H1): There is a tendency towards reciprocation in studied discussion networks ($i \rightarrow j$ and $j \rightarrow i$). (Dyadic Level)

Hypothesis 2 (H2): There is a tendency towards transitivity (i.e. increasing transitivity and reducing distance between actors; $i \rightarrow j$, $j \rightarrow k$ and $i \rightarrow k$). (Triadic Level)

Hypothesis 3 (H3): There is a tendency towards the increasing volume of interactions between learners themselves.

Hypothesis 4 (H4): There is a tendency towards preferential attachment within the studied networks.

PRELIMINARY RESULTS

Descriptive statistics of the discussion network

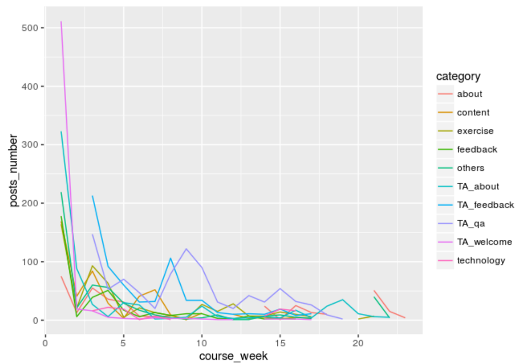
In 2014 MOOC, 1915 participants posted 5251 messages in total, of which 217 are threads, 5034 are replies and comments, while in 2015 psychology MOOC, 962 threads were provided, and 3097 are replies and comments.

In 2014 Psychology MOOC, there are topics initiated by TAs to collect feedbacks for individual sections and to answer content-related Q&A for each section. As shown in Figure 1, the number of the postings falling into the discussing categories initiated by TAs is relatively larger than the number of the same topics which are initiated by learners themselves. The category "content-related Q&A initiated by TAs for individual sessions" seems to attract a good number of replies and comments over time. Interestingly, as shown in Figure 1, the discussions of exercises share a similar quantitative pattern of content-related discussions; while the enquiries about the logistics of the course follow a similar pattern of technical discussions in both two offerings of psychology MOOCs. In

2015 Psychology MOOC, technical problems occurred during the mid-examination, showing as a peak in Figure 1.

Network Dynamics

Table 2 and 3 present the results of SIENA estimation. As shown in Table 2 and 3, the results of Model 0 (network effects: reciprocity; transitivity) indicate a tendency for participants to create mutual relationships at both dyadic and triadic levels, which leads to cohesiveness in the studied networks. This confirms that hypothesis H1 and H2 are accepted. The exceptional case is the transitivity effect identified in the category of "feedback" (i.e. general feedbacks to instructors and TAs initiated by learners), where there is no tendency for participants to create mutual relationship at triadic levels. This deserves a detailed examination in the future analysis. Interestingly, under the topic categories of "feedback" and "TA about" (i.e. enquiries about the logistics of the course initiated by TAs), when same role is used as a control variable, the transitivity effect is significant with a negative coefficient. Compared to discussions in other categories, it is less likely to create cohesive subgroups when learners provide feedbacks to the course and enquiries about course logistics.



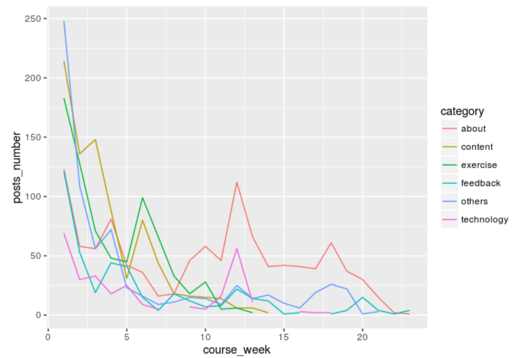


Figure 1. The number of postings within different discussion topics over time (2014 left & 2015 right).

In both courses, same role is a significant covariate effect with a negative coefficient. Thus, H3 (Model 1: reciprocity; transitivity; same role) is rejected, indicating that there is no tendency towards an increasing volume of interactions between learners.

H4 (Model 2: reciprocity; transitivity; Activity of alter) states that there is a tendency towards preferential attachment within the studied networks. The preferential attachment effect is not consistent among discussions of different topics. In most discussions, there is a tendency for participants who are actively involved in forum discussions in the early stages to become even more engaged over time. Nevertheless, when discussing exercises in 2014 Psychology MOOC, there is no preferential attachment effect, which deserves a future examination.

Category	Model 0	Model 1	Model 2
about	3.49* (0.37) 1.06* (0.29)	3.27* (0.40) 0.82* (0.32) -2.25* (0.24)	3.89* (0.35) 1.32* (0.31) -0.51 (0.34)
content	4.53* (0.32) 0.76* (0.23)	4.48* (0.32) 0.84* (0.23) -1.56* (0.28)	4.27* (0.29) 0.63* (0.23) 0.20* (0.09)
exercise	4.27* (0.35) 0.35 (0.34)	4.17* (0.34) 0.33 (0.36) -2.34* (0.35)	5.44* (0.46) 0.97 (0.43) -1.26* (0.60)
feedback	3.36* (0.50) -0.92* (0.41)	3.33* (0.50) -0.96* (0.40) -1.16* (0.44)	4.03* (0.78) -0.38 (0.74) -0.89 (0.93)
technology	3.26* (0.61) 0.03 (0.57)	3.02* (0.67) -0.15 (0.59) -2.18* (0.30)	3.63* (0.54) 0.36 (0.64) -0.53 (0.52)
TA about	5.13* (0.33) 0.16 (0.12)	3.67* (1.02) -0.41* (0.13) -5.91* (0.10)	4.71* (0.78) -0.81* (0.15) 0.20* (0.01)
TA feedback	3.30* (0.33) 0.87* (0.18)	0.64 (0.35) 0.08 (0.09) -4.98* (0.11)	0.67 (0.44) 0.28 (0.16) 0.12* (0.004)
TA Q&A	1.56* (0.44) 1.35* (0.17)	0.37 (0.48) 0.50 (0.08) -4.05* (0.10)	0.42 (0.46) 0.89* (0.17) 0.19* (0.01)

Table 1. Estimation results of network effects with standard errors in parentheses (2014 Psychology)

Category	Model 0	Model 1	Model 2
about	3.63* (0.21) 1.61* (0.17)	3.39* (0.23) 1.21* (0.17) -3.02* (0.19)	3.08* (0.22) 1.02* (0.20) 0.16* (0.01)
content	4.23* (0.23) 1.37* (0.20)	4.26* (0.22) 1.35* (0.20) -1.39* (0.49)	3.89* (0.26) 0.71* (0.21) 0.09* (0.01)
exercise	3.46* (0.27) 1.23* (0.23)	3.52* (0.26) 1.23* (0.24) -0.02 (1.18)	3.28* (0.25) 1.04* (0.23) 0.11* (0.04)
feedback	3.68* (0.37) 1.03* (0.33)	3.50* (0.38) 1.01* (0.34) -2.69* (0.28)	3.33* (0.35) 0.82* (0.37) 0.28* (0.10)

Table 2. Estimation results of network effects with standard errors in parentheses (2015 Psychology)