Supporting the Encouragement of Forum Participation

Aashna Garg  
Stanford University  
aashna94@stanford.edu

Andreas Paepcke  
Stanford University  
paepcke@cs.stanford.edu

ABSTRACT
We analyze the forum contribution rates during forty offerings of 12 college courses reaching back to 2011. The goal is to identify moments during the quarter when encouragement for passive students might be timely. We started with a social graph model of each forum archive. From those models we computed weighted out degrees, and page ranks for each student. The out degrees reflect the number of posts a student contributes. We performed change point analyses through bootstrap procedures over the CUSUM data of these post contributions through each quarter. We thereby identified significant week by week changes in the rate at which the top ten percent of forum contributors post messages. We hypothesize that such change points might be appropriate encouragement opportunities, and we find that sudden rate shifts do occur along regular patterns, primarily in science and engineering courses. We propose and demonstrate the use of control charts to monitor forum traffic, and show how the historic data can be used to provide personalized encouragement messages.

1. INTRODUCTION
Many massively open online course (MOOCs) rely heavily on online forum facilities. The remote course participants are expected to communicate by posting questions and opinions on the learning management system’s forum. Instructors and teaching assistants sometimes selectively respond via that medium. Importantly, the courses rely on the students themselves to answer each others’ questions in comments to posts. Whereas in science courses the forum often serves primarily the exchange of questions and answers, humanities courses frequently require discussions among students. The forum plays an even more central part in those scenarios.

Given their importance in MOOCs, a number of studies have examined student behaviors on MOOC forum facilities. For example, forum participation has been studied as an early indicator of student dropout [17]. The distribution of forum post submissions among course participants was examined in [10]. This examination showed that frequent posters, or superposters do better in courses than the average among other participants. The authors further find that the presence of superposters has positive impact on the rest of the forum participants. The importance of actively fostering interactions was shown in [?].

But forum facilities have also steadily gained acceptance in residential, classroom-bound courses. The forum is a supplemental to in-class discussion—if such face-to-face interactions even occur in the lecture hall. Studies on how academic outcomes are related to forum participation in residential settings are missing because access to grades in these settings tends to be highly restricted. This work suffers from this limitation as well.

However, we can at least hypothesize that benefits such as those found by [10] do carry over to residential students. If we accept for the moment that forum participation is indeed beneficial, then we need to ask how, and when instructors can encourage students to participate with posts and comments on others’ contributions.

In a MOOC setting, [11] showed that variously framed emails encouraging the use of the forum yielded mixed results. But residential students operate under very different incentive structures. Most desire a successful completion of each course. Even large residential classes are much smaller than MOOCs, and students often know at least some of their peers from face-to-face meetings. We might therefore expect that reactions to encouragement might differ from those in MOOC contexts.

We analyzed a number of Piazza [14] forum posting behaviors in a large residential college. Our goal with this preliminary work was to find promising points during a quarter for when to send encouraging emails to students who are overly passive on the forum. We also propose a data-informed method for personalizing such messages.

2. RELATED WORK
Many educators expend enormous amounts of effort to design their courses towards maximizing the value of student interactions. Regardless of the approach taken, a series of questions consistently arises: How effective is the course? Is it meeting the needs of the students? How can the needs of learners be better supported? What interactions are effective? How can they be further improved? Traditional approaches to answering these questions have involved student evaluation, the analysis of grades and attrition rates, and instructor perceptions most often gathered at the end of a course [7].

Norris et al. [4] suggest that they would like analytics to be used to measure, compare and improve the performance of individuals, not just to better the experience but also to facilitate better outcomes to the activity.
Dawson et al. [5] added that the analysis of educational data might be used to improve the student learning experience. Such analysis would quantitative as well as qualitative approaches. The interpretation would have to include a value judgment on people’s use of the environment: not only counting who uses the environment for what, but also judging what might be a good and what might be a bad experience, and offering suggestions for moving on the continuum from one to the other.

Positive learning behaviors can be encouraged through the use of persuasive technologies. Any knowledge about given learners’ behaviors, complemented with those of their peers, and the assumed theoretical ideal, could be used as a nudge. It is argued that if triggered by contextual clues, positive “nudges” may help students achieve their learning goals [16]. Fogg [8] anticipated that in the future students could be nudged in exactly this manner towards learning success.

However the successful application of these technologies presupposes a very good understanding of the learners’ behavior. Whilst such an understanding can be sought through learning analytics, this is a challenging endeavour as the data required may be incomplete, inaccurate, and technically difficult to both collect and process in real-time [16].

3. DATASET
Since 2011 many courses in the studied university have used the Piazza forum facility for its internal courses. Our dataset comprises the posts and comments of these usage years. From among 5000 university course offerings that have used the Piazza forum facility since 2011 we selected twelve courses, with a total of around 40 offerings. We use the term offering to denote a course taught during a particular quarter. We call the set of all a course’s offerings a course. Most courses had multiple offerings since 2011, so we were able to include longitudinal forum usage data for most courses.

We used two criteria for selecting the courses to analyze. We favored those that had comparatively large numbers of student posts, and we tried to cover courses from many schools and departments. We in particular sought to include Humanities courses that made use of Piazza for class discussion. Figure 1 summarizes our choice. Figure 2 communicates an impression of relative post numbers across offerings (neighboring swatches), and courses.

From the forum posts of these data we constructed one social graph for each course offering. We then analyzed changes in two social graph measures over time to find ‘important’ weeks.

4. FROM POSTS TO CONNECTION GRAPH
Social networks are most simply modeled by considering each participant as a node, and interactions initiated by participants as out-directed links. In our case each node represents a student, i.e. we only define one type of node. Links in our forum interaction graph represent one student posting, or comment on another student’s contribution. That is, our links are unidirectional. Multiple interaction initiations by one student are captured by weighting the corresponding outgoing links. Many graph analysis tools operate on models of this type.
For the purpose of identifying candidate time points for encouraging online conversation participation our chosen model suffices. Note that this work does not consider additional measures, such as content quality, such as irrelevant postings, for which a richer model would be required.

Many measures are used to quantify various aspects of social graphs [9, 12]. Not all are meaningful in the context of education-related forum interactions. We focus here on two measures: weighted out-degree, and page rank. Figure 3 illustrates our use of social graphs for forum activity. Nodes A, B, C, D, and E represent students. The link from A to B is marked with the number 3, because A commented three times on one or more of B’s posts. The number of outgoing links is the node’s out-degree weight. For example, the out-degree weight of A is 4.

The number of incoming links is called the node’s in-degree. Node C’s in-degree is 1. Node E has no links entering or exiting. The respective student either did not participate in the forum or never replied to others’ posts and never received replies on their posts.

Analogous to Web pages, each node can be assigned a page rank. The intuition in this context is that student S₁’s presence in the forum is more ‘important’ than student S₂’s if the node representing S₁ has higher page rank than the node that represents S₂. In our context the intuition behind page rank is that a node N is more important (has higher page rank) the more other important nodes comment on N’s posts. Imagine a scenario in which student S₁ posts an interesting question, to which many students comment with their opinion, creating a long thread. The node representing S₁ would experience an increase of its page rank with every incoming comment. Node B in Figure 3 is an example for this situation. Its in-degree weight is 4. If B were to comment on one of D’s posts, then D’s page rank would increase more than if the low-page-rank node C commented on D.

In terms of evaluating a student’s participation in the forum, a high page rank, and high out-degree are positive. Low values are less positive. We chose these two measures because of their relatively straight-forward meanings when applied to forum posts, and for their relevance to our goal of identifying potential intervention times. Some of the fifteen other measures we computed, such as betweenness are meaningful for forum scenarios as well, but their usefulness depends on one’s analysis goals. For example, [17] include several of those measures for the purpose of persistence prediction. For evaluating contribution quality the contents of posts would need to be considered: students who persistently post irrelevant contents contribute less positively to the forum than constructively participating students. However, for our purposes the two measures of page rank and out-degree provide strong enough signals. Because of space constraints we focus in the following material on the weighted out-degree to explore its contributions to insights into student behavior.

5. ANALYSIS PROCEDURE

For each course we identified the top 10% of students who contributed the most to the forum. For those students we computed the average of our chosen social graph measures. This average was computed for each week of each offering. The results provided a week-by-week time series of postings and evolving pagerank for these students.

Figure 4 shows a typical time series with the sum of postings (i.e. weighted out degree) on the ordinal, and weeks into the quarter on the abscissa. Notice in Figure 4 that the successive increases in average contribution of the top 10% appear to take a jump after week 5. Such a discontinuity, or change point could be a good trigger for encouraging students who have held back participation. For example, the jump could be due to important discussions just before the midterm.

However, visual inspection is insufficient for determining whether this change is statistically significant, or could instead be explained by random noise. The question mark in the figure is a reminder of this uncertainty.

Therefore, we next performed an analysis aimed at quantifying how much a given increase or decrease in weekly posts differed from the mean of those variations. For each series of course offerings this analysis yielded insight into how consistent the location of change points was across multiple offerings of the same course. Such consistency across offerings would imply that an instructor could ‘hard code’ the
In order to discover for each week whether its number of contributions differed significantly from those of other weeks, we performed a bootstrap procedure across the post-contribution increments, i.e. over the data points of the top line in Figure 5 [15] This process takes as reference the CUSUM (cumulative sum) over the incremental post counts [6]. This quantity is computed for each week as the sum of differences from the overall mean of incremental posts.

\[ S_m = \sum_{i=0}^{m} (x_i - \bar{\mu}_0) \]  

(1)

where \( x_i \) is the \( i \)th weekly post increment, and \( \bar{\mu}_0 \) is the mean across all the weekly post increments.

Our bootstrap permuted the post increments, and re-computed the CUSUM points 1000 times. For each permutation we computed the \( S^0_{\text{diff}} = S^0_{\text{max}} - S^0_{\text{min}} \) of the resulting CUSUM, and determined whether \( S^0_{\text{diff}} \) was less than the corresponding value \( S_{\text{diff}} \) in the CUSUM of the correctly ordered number of posts. \( S^0_{\text{max}} \) and \( S^0_{\text{min}} \) are the maximum/minimum values of the CUSUM for a given iteration. Their difference represents the farthest that the iteration’s CUSUM strays from the CUSUM’s mean.

We compute a confidence \( C \) that a change point occurred sometime in a quarter via:

\[ x_i = \begin{cases} 1, & \text{for } S^0_{\text{diff}} < S_{\text{diff}} \\ 0, & \text{for } S^0_{\text{diff}} \geq S_{\text{diff}} \end{cases} \]  

(2)

\[ C = 100 \times \frac{1}{N} \sum_{i=1}^{N} x_i \]  

(3)

where \( x_i \) is the outcome of one bootstrap iteration, and \( N \) is the number of bootstrap iterations. Figure 6 shows an example of five iterations. Points on the unlabeled lines denote differences of CUSUM points from the mean during one week. Note that the distance of those lines from the dotted mean line tends to be much smaller than the CUSUM distance for the correctly sequenced values at the bottom. Following [15] we determined the last week before a change for each change point. Figure 5 shows an example.

The bottom chart’s line shows again the weekly post contributions. Up to about week four, weekly postings by the top-10% was below average. The corresponding downward segment of the CUSUM reflects increasingly negative sums of those differences. During week 5 weekly contributions begin to rise above the mean, and the CUSUM ‘recovers’ upward towards zero.

The horizontal (red) lines in the top chart indicate the upper and lower limits within which week-to-week differences in posts contributions would be expected to vary just by chance\(^1\). The lines are called control limits, and we will discuss their potential real-time use during a course’s run time in Section 6.

The vertical bands in Figure 5 (bottom) indicate visually where a significant change in contribution rate occurred. In this case the sharp increase at week five was found via Equation 3 to be a true change point at confidence level 92%.

\(^1\)Assuming normal distribution of contribution difference values
6. RESULTS AND DISCUSSION

Our question was “when would be good times during a course for encouraging students who lag behind in forum contributions?” We hypothesize that times when the top 10% of contributors speed up, other students might be encouraged to do the same.

Each row in Table 1 shows to which degree offerings of that row’s course over the years showed consistent changes in behaviors of the top-10% contributors. The percentages in a week’s column indicate the proportion of the respective course’s offerings that experienced jumps during that week. The results indicate that week six is particularly likely to experience changes in posting rates. Given that the university operates on a quarters schedule, we can hypothesize that the traffic is related to midterms. Week eight might be related to final projects coming due not too far out. Note, however, that the number of offerings (shown in parentheses with the course names) differ across courses. Thus 20% for urban studies means only one of five offerings experienced a significant posting acceleration.

The only courses with an appreciable number of change points among their offerings are engineering courses. The humanities and social sciences, while using forum facilities, have not included the forum as a central discussion hub. The numbers of students attending these courses are also smaller than the science/engineering classes.

When computing the percentages of Table 1 we included all offerings, reaching back to 2011. This year was one of the earliest that Piazza offered forum services. It is possible that as forum use continues to gain momentum, more regular patterns might emerge.

The forum change point computations we outlined above require data from all weeks of an offering to be available. In the presence of historic data this requirement is not a problem. But what could an instructor do to discover unusual posting frequencies while the offering is running?

6.1 Control Charts for Forum Alerts

Figure 5’s lower chart hints at a possibility we have not yet discussed. As mentioned, the horizontal dotted lines are process limits, a term from process control practice [6]. The limits bound the values within which a process is expected to vary. For industrial processes the variations might lie around a known optimal operating level, such as a temperature. When no such level is known from the domain, a mean can be used. Our process limits denote \(2 \times \sigma\) distance from the mean.

In the context of instructional forum use it would be possible to detect points, such as the one above the upper control limit in Figure 5. Such change to above-normal might indicate confusion among the students, or the discovery of an exciting topic. Either way, the instructor’s attention might be warranted, as might be pointing passive students to the increased activity.

6.2 Personalized, Quantitative Encouragement

The data framework we discussed could also be deployed to provide personalized encouragement for passive students.

Figure 7: Personalized, quantitative nudges: catching up with the top-10% within five weeks, or two weeks.

Rather than admonishing students for past passivity, a forward-looking nudge could be provided. Figure 7 illustrates this option. The dot represents one student and their contribution as of week six: two postings. Two of many possible options are shown in the Figure for catching up to the top-10% contributors. These, and other options are derived by drawing a line from the student’s position to an intersection with the regression line. That line is known from past course offerings. The larger the slope of the connecting line, the more aggressive the plan for catching up. For example, the long line would catch the student up within five weeks, assuming a weekly rate of \((20 - 2)/5 \approx 4\) messages per week.

Alternatively, the shorter line would call for six messages per week, to catch up within two weeks. The square at the end of the long option is intended to represent a slider that the student could run along the regression line to make a plan. The number of required weekly messages would be updated continuously as the student operates the slider. While this sketch is certainly not the ideal, and ultimate user interface, it illustrates the ideas of using past and current forum data to provide (i) forward-looking encouragement in place of retribution, and (ii) to empower the student by personalizing the message, and providing a tool for planning. Studies are needed to determine whether postings in response to even such personalized messages prove beneficial.

Yeomans et al. [19] suggest that planning prompts can help learners adopt productive frames of mind at the outset of a learning goal that encourage and forecast student success.

7. CONCLUSION AND FUTURE WORK

We studied how forum posting rates vary week to week across about 40 offerings of courses. We included science, engineering, music, political science, urban studies courses, and others. The purpose was an investigation whether particular weeks in each quarter often stand out as featuring significant changes in post rates. We found that such weeks exist, albeit not with perfect consistency, and mostly lim-
The important next step is to test whether encouragements at the change points are in fact more effective than encouragement offered, for example, at the beginning of a course. Equally important is to extend our study to include measures of post quality. We focused on the top 10% most prolific forum contributors. But the quality of those postings—by some measure—is important to include in evaluations of forum health. For that exploration we will include both page content and the page ranks we computed for this work. Our graph-based approach has generated the necessary framework for this future investigation.

A further use of page rank will be to observe how teaching assistants interact with students around particular topics. TA participation has been filtered out from the our study to prevent bias. TAs will tend to have high page rank, and by filtering page rank computations by the topics of student postings we can expect to see topic-based divisions of labor among the assistants. They can then more formally distribute their load.

Forum facilities offer potential for student interaction outside the classroom. One hopes that courses with a large component of discussion might benefit in the future. Meanwhile, it is important to develop methods that help all students take advantage of these online contact options. This work is a step in that direction.

Table 1: Consistency of behavior changes across offerings in weeks with change points.

<table>
<thead>
<tr>
<th>Course</th>
<th>Week3</th>
<th>Week4</th>
<th>Week5</th>
<th>Week6</th>
<th>Week7</th>
<th>Week8</th>
<th>Week9</th>
<th>Week10</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI Principles (6)</td>
<td>17%</td>
<td></td>
<td></td>
<td>50%</td>
<td>16%</td>
<td>33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Machine Learning (7)</td>
<td>50%</td>
<td>14%</td>
<td>14%</td>
<td>14%</td>
<td>28%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Vision (4)</td>
<td></td>
<td>50%</td>
<td></td>
<td></td>
<td></td>
<td>50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision Analysis (2)</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td></td>
<td></td>
<td></td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td>Algorithms for</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>Molecular Biology (6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>International</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urbanization (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8. REFERENCES