A Stackelberg game approach for incentivizing participation in online educational forums with heterogeneous student population

Rohith D. Vallam and **Priyanka Bhatt** and **Debmalya Mandal** and **Y. Narahari**
Department of Computer Science and Automation,
Indian Institute of Science, Bengaluru, 560012, India
{rohithdv, priyanka.bhatt, hari}@csa.iisc.ernet.in, debmalya.mandal@gmail.com

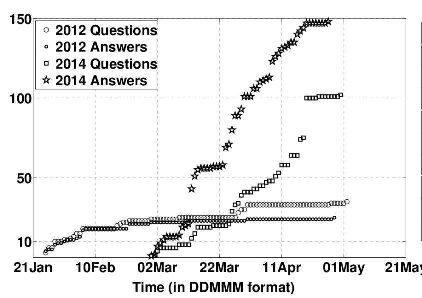
Abstract

Increased interest in web-based education has spurred the proliferation of online learning environments. However, these platforms suffer from high dropout rates due to lack of sustained motivation among the students taking the course. In an effort to address this problem, we propose an incentive-based, instructor-driven approach to orchestrate the interactions in online educational forums (OEFs). Our approach takes into account the heterogeneity in skills among the students as well as the limited budget available to the instructor. We first analytically model OEFs in a non-strategic setting using ideas from lumpable continuous time Markov chains and compute expected aggregate transient net-rewards for the instructor and the students. We next consider a strategic setting where we use the rewards computed above to set up a mixed-integer linear program which views an OEF as a single-leader-multiple-followers Stackelberg game and recommends an optimal plan to the instructor for maximizing student participation. Our experimental results reveal several interesting phenomena including a striking non-monotonicity in the level of participation of students vis-a-vis the instructor's arrival rate.

Introduction

With the explosive growth of the Internet, the area of education has undergone a massive transformation in terms of how students and instructors interact in a classroom. Online learning environments now constitute a very important part of any academic course. Further, online education has attracted the interest of the research community due to the immense popularity of the massive open online courses (MOOCs) offered by platforms like Coursera, edX, Udacity, etc. As of January 17, 2014, Coursera students voiced themselves in 590,000 discussion threads in the education forums for a total of 343,014,912 minutes of learning across 571 courses (Coursera 2014). However, empirical studies have repeatedly shown that the dropout rates in the online courses are very high (Fowler 2013) mainly due to a lack of sustained motivation among the enrolled students. An important, but often under-utilized component of an online classroom is the online educational forum (OEF) where students and instructors discuss various administrative and technical aspects of the course (Andresen 2009; Mazzolini and Maddison 2003). The objective of our work is to propose an instructor-driven approach to orchestrate the activities of OEFs by designing optimal incentives to enhance student-instructor participation in these OEFs.

Incentive design plays an important role in encouraging participation among students in these educational forums. As part of a case study, we analysed the data collected from two online educational forums which were part of the Game Theory (E1 254) in the Department of Computer Science and Automation, Indian Institute of Science, for two different terms. The primary difference between these two terms was that there were no incentives offered to students participating in the Spring 2012 term while in the Spring 2014 term, students were offered incentives (a certain percentage of marks based on the reward points accumulated by the student) to participate actively to open-ended (or discussion type) questions posted on the Piazza forum associated with the course. We observed an increased participation of students in the incentive-based course than when there were no incentives offered (See Figure 1) which is an indication of importance of appropriate incentives in driving up the participation levels in the course.



OEF Details		
Course Term	Class Strength	Forum (Piazza) Duration
2012	72	Jan 21 to May 1
2014	38	Mar 1 to May 1

Figure 1: Impact of incentives on participation of students in OEFs recorded in a real-world experiment.

Modelling incentives for improving the participation levels has also recently been studied by (Ghosh and Kleinberg 2013) where they formulate an incentive-based approach to modulate the activities of the online educational forums and recommend optimal behaviour for the instructor needed to drive up the participation from the student population. We pursue this direction further and non-trivially extend their instructor-driven approach to a more realistic setting of heterogeneous students and a budget-limited instructor.

Online learning environments attract participation from students with heterogeneous skill levels and it is our belief that any approach to improve levels in participation should account for this heterogeneity which is not considered by existing models. We incentivize students on a per question basis to keep up the momentum of participation in the class. Students are provided suitable incentives to post answers to the specific open-ended/discussion-style questions that are posted on the forum by the instructor. These incentives maybe in the form of some book vouchers, food coupons or some extra grade points, as considered appropriate by the instructor. The instructor is limited by a budget and has to make judicious use of them such that it results in higher participation levels from different types of students.

Contributions and Outline

Our work requires us to first define an interaction model in an OEF which reasonably captures the activities of the students and instructor. Once this is addressed, there is a need to understand transient behaviour of the instructor and students in the time-limited online course. One of the complexities is to handle the 'continuous' nature of arrivals of the instructor and students to the OEF. Taking into consideration these factors, we model the OEF as a continuous-time Markov chain (CTMC). A CTMC is a natural candidate to describe the interaction model as it enables us to record the activities of all the students and instructor in a simple and elegant manner by appropriately modelling the state space of the CTMC. Further, it allows the state transitions to happen at any epoch of time. Using techniques from lumpability of CTMCs, we compute the transient behaviour of the instructor and the students in the modelled OEF.

Next, we use these computations in a more realistic game-theoretic setting where we adopt a Stackelberg game approach to model the interactions of the OEF. Stackelberg game models (for example: (An et al. 2011)) have been the natural approaches in many well-known practical applications primarily in the context of security (for example: for deploying surveillance resources (Jain et al. 2010) by the Los Angeles International Airport (LAX) police) and more recently, in traffic patrolling for the Singapore road network (Brown et al. 2014). We believe that our problem, though it is set in a different domain of education, fits naturally into a Stackelberg framework where the players of the game compete on a resource (i.e., participation time on the OEF) and the welfare maximizing leader (i.e., instructor) is in a position to exploit the first-mover advantage to trigger increased participation from the followers (i.e., student population) by designing suitable incentive schemes. Further, our detailed experiments with the proposed Stackelberg model demonstrate that our approach validates several empirically/theoretically observed phenomena and also, offers utility-maximizing recommendations to the instructor as well as the different types of students on several important parameters (like arrival rate, instructor bias) of the model.

In the rest of the paper, due to space constraints, we sometimes omit providing details in the interest of clarity. These omitted details are given in (Vallam et al. 2014).

```
Instructor chooses an arrival rate to the forum and announces this to the class.
Students (of different types) observe the instructor announcement and decide their
  corresponding rate of arrival.
Instructor and the student record the next arrival time based on their corresponding
  chosen arrival rates.
while Course has not ended do
    if you are the instructor then
        if there is time available for next arrival then
            Engage in other activities not related to the forum.
        else
            Close current (open-ended) question on the forum.
            Reward points to students who have answered this question.
            Post the next discussion-style question.
            Record the next time for arrival to the forum.
        end if
    end if
    if you are a student then
        if there is time available for next arrival then
            Engage in any other activities not related to the forum.
        else
            Post a valid answer to the current open question in the OEF.
            Record the next time for arrival to the forum.
        end if
    end if
end while
```

Figure 2: An instructor-driven interaction model in an OEF

Problem Setting

We consider n students and an instructor participating in an hybrid or online classroom which has an associated discussion forum (termed in this paper as online educational forum (OEF)). In the rest of the paper, we focus on modelling the instructor-student interactions on the OEF. To achieve the right focus, we do not consider any other aspect of the online classroom like lectures, written assignments, exams, etc. We assume that the arrivals of the n students and the instructor to the OEF are independent Poisson processes with rate λ_i for student i ($\forall i \in \{1, \ldots, n\}$) and rate μ for the instructor.

We begin by proposing an instructor-driven approach to structure the activities in the OEFs (see Figure 2). Henceforth, we will assume the activities of the OEF follows as outlined in Figure 2. We capture the heterogeneity among students by allowing L types of students in our OEF. Let A_l be the set of all students of type l ($\forall l \in \{1, \ldots, L\}$). Also, let n_l be the number of students of type l. On arrival to the forum, a student of type l answers the currently open question (if any) and hence, incurs a cost (α_l). We assume that the cost of answering a question will be same for all students of a given type. The instructor has a budget B per question and has to decide a suitable allocation of the budget among the different students belonging to the L types. Let m_l denote the maximum number of answers *per question* that can be given to A_l students i.e., $B = \sum_{1 \leq l \leq L} m_l$.

A CTMC Model for OEFs

We now model the activities of the OEF (as described in Figure 2) as a CTMC $X(t) = (\mathcal{S}, Q)$, where $\mathcal{S}$ is the set of states of the stochastic process $X(t)$ and Q is the generator matrix. We define each state $x \in \mathcal{S}$ as: $x = (x_1, \ldots, x_n)$, where x_i corresponds to the number of answers received from the i^{th} student for the current question. If $X(t)$ is in state $x : (x_1, .., x_i, .., x_n)$ and student i gives an answer on the forum then $X(t)$ transitions to state $y : (x_1, .., x_i + 1, .., x_n)$. When the instructor arrives at the forum, the current question is closed and a new question is started, thus transitioning to state $(0, \ldots, 0)$. At any point of time, there is a single active question on the OEF as given in Figure 2. We assume the course is of finite duration (T) and the instructor and the students are arriving to the forum at finite rates. This allows us to obtain a reasonably large upper bound M on the number of answers that a student can post on the forum for each question transforming our infinite state CTMC to a finite state CTMC where the 'last' state will be $(M, .., M)$. Figure 3(a) illustrates the CTMC for two students. Thus, the generator matrix of the CTMC can be defined as below.

$$Q(x,y) = \begin{cases} \lambda_j & \text{if } \sum_i |x_i - y_i| = 1 \text{ and } \exists j : y_j - x_j = 1 \\ \mu & \text{if } y = (0, \ldots, 0) \neq x \\ \omega & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$$

$$\text{where } x, y \in \mathcal{S}, \text{ and } \omega = \sum_{y' \in \mathcal{S} \setminus \{x\}} -Q(x, y')$$

Lumpability of the OEF CTMC

An important point to note is that the rewards and costs incurred for each student in a particular state are independent of the other students and dependent only on their own arrival rates and the instructor's arrival rate. This is possible because only open-ended questions are being posted by the instructor and thus, even if an open-ended question has already been answered by a few students, still a new student can find it beneficial to give a new answer and potentially earn a good reward. Hence, instead of analyzing the CTMC, which keeps track of arrivals of all students (of all types), we show that we can analyze n independent student-specific CTMCs (with $M+1$ states each) so that each of the student-specific CTMC keeps track of arrivals from only that particular student. This is possible by applying the lumping process on the original CTMC which we describe next.

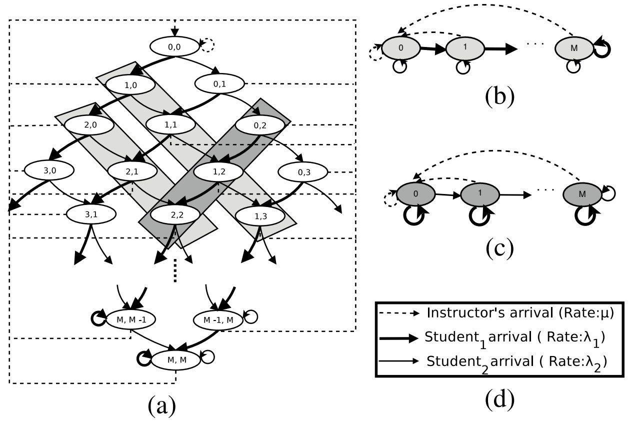


Figure 3: (a) Light-gray and dark-gray regions depict two partitions that can be defined on the CTMC. Each light-gray region contains states where the number of arrivals of student 1 is the same. Each dark-gray region contains the states where the number of arrivals of student 2 is the same. (b) Lumped-Student$_1$ CTMC: Each light-gray state indicates the aggregation of all states in the Figure 3(a) which are enclosed by light-gray region. For example, state 1 denotes the set of states $\{(1,0),(1,1),\ldots,(1,M)\}$ from Figure 3(a). (c) Lumped-Student$_2$ CTMC: Each dark-gray state indicates the aggregation of all the states in the Figure 3(a) which are enclosed by dark-gray region. For example, state 2 denotes the set of states $\{(0,2),(1,2),\ldots,(M,2)\}$ from Figure 3(a). (d) This is the legend for all the diagrams.

We first define a partition $\overline{S}_i$ on the state space $\mathcal{S}$ of $X(t)$ w.r.t. a student i in the OEF as $\overline{S}_i = \{\overline{S}_i^a | a \in \{0, 1, \ldots, M\}\}$ where each block $\overline{S}_i^a$ of the partition $\overline{S}_i$ is defined as $\overline{S}_i^a = \{(x_i, x_{-i}) \in \mathcal{S} | x_i = a\}$. Figure 3 depicts the lumping process in more detail through an example. We now state and prove a result about the lumpability of the proposed CTMC into smaller, student-specific CTMCs. For ease of notation, note that all the notations with an overline refer quantities corresponding to the lumped CTMC.

Theorem 1. ***(i)*** *$X(t) = (\mathcal{S}, Q)$ is lumpable w.r.t. partition $\overline{S}_i = \{\overline{S}_i^a | a \in \{0, 1, \ldots, M\}\}$.*

(ii) *The quotient (lumped) Markov chain $\overline{X}_i(t) = (\overline{S}_i, \overline{Q}_i)$ that we get on lumping the CTMC $X(t)$ w.r.t. partition $\overline{S}_i$ ($i \in \{1, 2, \ldots, n\}$) is given as :*

$$\overline{Q}_i(\overline{S}_i^a, \overline{S}_i^b) = \begin{cases} \lambda_i & b = a+1, \\ \mu & b = 0 \neq a, \\ \omega & b = a, \\ 0 & o/w. \end{cases}$$

$$\textit{where, } \omega = -\sum_{c \in D \setminus \{a\}} \overline{Q}_i(\overline{S}_i^a, \overline{S}_i^c), D = \{0, 1, \ldots, M\}.$$

Proof. *(Sketch)* (i) $X(t)$ is lumpable w.r.t $\overline{S}_i$ if for any two blocks $\overline{S}_i^a, \overline{S}_i^b \in \overline{S}_i$ and for every $v, y \in \overline{S}_i^a$ we have $q(v, \overline{S}_i^b) = q(y, \overline{S}_i^b)$ i.e. the rate of transition from each state in block $\overline{S}_i^a$ to block $\overline{S}_i^b$ should be equal. By definition, $\overline{S}_i^a = \{(x_i, x_{-i}) \in \mathcal{S} | x_i = a\}$ and $\overline{S}_i^b = \{(x_i, x_{-i}) \in \mathcal{S} | x_i = b\}$. Now states $v, y \in \overline{S}_i^a$, hence these maybe represented as: $v = (a, v_{-i})$ and $y = (a, y_{-i})$ respectively. Also we represent a state $z \in \overline{S}_i^b$ as: $z = (b, z_{-i})$. Let $D = \{0, 1, \ldots, M\}$ and $\mathcal{S}_{-i} = \{(x_{-i}) | (x_i, x_{-i}) \in \mathcal{S}\}$. Now we need to prove that $\forall \overline{S}_i^a, \overline{S}_i^b \in \overline{S}_i$ and for any $v, y \in \overline{S}_i^a$,

$$q(v, \overline{S}_i^b) = q(y, \overline{S}_i^b) \quad \Rightarrow \sum_{z \in \overline{S}_i^b} Q(v, z) = \sum_{z \in \overline{S}_i^b} Q(y, z)$$

$$\Rightarrow \underbrace{\sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, v_{-i}), (b, z_{-i}))}_{\textbf{LHS}} = \underbrace{\sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, y_{-i}), (b, z_{-i}))}_{\textbf{RHS}} \tag{1}$$

- Case (1) $b \neq a$
  - Case (1a) $b = a + 1$. It can be shown that **LHS** $= \lambda_i =$ **RHS**.
  - Again, it can be shown **LHS** $=$ **RHS** $= 0$ for scenarios: Case (1b) $b > a + 1$, Case (1c) $b < a, b \neq 0$
  - Case (1d) $b = 0, b \neq a$. It can be shown that **LHS** $= \mu =$ **RHS**.
- Case (2) $b = a$. Using this and simplifying Equation (1), we get

$$\textbf{LHS} = -\sum_{(z_i \in D \setminus \{a\})} \sum_{(z_{-i} \in \mathcal{S}_{-i})} Q((a, v_{-i}), (z_i, z_{-i}))$$

$$\textbf{RHS} = -\sum_{(z_i \in D \setminus \{a\})} \sum_{(z_{-i} \in \mathcal{S}_{-i})} Q((a, y_{-i}), (z_i, z_{-i}))$$

Now, from Case (1) we have, $\forall c \neq a$,

$$\sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, v_{-i}), (c, z_{-i})) = \sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, y_{-i}), (c, z_{-i})),$$

$$\Rightarrow \sum_{(z_i \in D \setminus \{a\})} \sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, v_{-i}), (z_i, z_{-i})) =$$

$$\sum_{(z_i \in D \setminus \{a\})} \sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, y_{-i}), (z_i, z_{-i})). \therefore \textbf{LHS = RHS}$$

(ii) The generator matrix for the lumped CTMC $\overline{X}_i(t) = (\overline{S}_i, \overline{Q}_i)$ can be got by $\overline{Q}_i(\overline{S}_i^a, \overline{S}_i^b) = q(x, \overline{S}_i^b) = \sum_{z \in \overline{S}_i^b} Q(x, z)$ for any

$\overline{S}_i^a, \overline{S}_i^b \in \overline{S}_i$, where $x \in \overline{S}_i^a$. These quantities have been computed in Cases (1) and (2) above. $\square$

We now have n lumped-CTMCs $\overline{X}_i(t) = (\overline{S}_i, \overline{Q}_i)$, $1 \leq i \leq n$ (See Figure 3(b) and Figure 3(c)) with finite state space $\overline{S}_i = \{\overline{x} | \overline{x} \in \{0, 1, \dots, M\}\}$. Each block $\overline{S}_i^{\overline{x}} \in \overline{S}_i$ has been represented as a state $\overline{x}$ of the lumped-CTMC $\overline{X}_i(t)$ i.e. a state $\overline{x}$ of the lumped-CTMC $\overline{X}_i(t)$ is representative of the block $\overline{S}_i^{\overline{x}}$ which contains all the states in $\mathcal{S}$ in which the student i arrives $\overline{x}$ number of times. Each state $\overline{x} \in \overline{S}_i$ thus simply means how many answers have been received from student i. For notational ease, we sometimes denote $\overline{S}_i^a$ as $\overline{a}$. Also, let $\overline{\pi}_i^t$ denote the transient state probability vector for the lumped CTMC $\overline{X}_i(t)$. Transient state probability $\overline{\pi}_i^t(\overline{x})$ is the probability of $\overline{X}_i(t)$ being in the state $\overline{x}$ at a time instant t. We need to solve the following differential equations for the transient probability vectors of $\overline{X}_i(t)$:

$$\frac{d\overline{\pi}_i^t(0)}{dt} = -\lambda_i \overline{\pi}_i^t(0) + \sum_{\overline{y}=1}^{M} \mu \overline{\pi}_i^t(\overline{y}); \quad \frac{d\overline{\pi}_i^t(M)}{dt} = -\mu \overline{\pi}_i^t(M) + \lambda_i \overline{\pi}_i^t(M-1)$$

$$\frac{d\overline{\pi}_i^t(\overline{x})}{dt} = -(\lambda_i + \mu)\overline{\pi}_i^t(\overline{x}) + \lambda_i \overline{\pi}_i^t(\overline{x}-1), \forall \overline{x} : 0 < \overline{x} < M$$

The initial state distribution $\overline{\pi}_i^0$ for the CTMC $\overline{X}_i(t)$ is $\overline{\pi}_i^0(0) = 1$ and $\overline{\pi}_i^0(\overline{x}) = 0 \forall \overline{x} \in \overline{S}_i \setminus \{0\}$ as, initially, no answer would be posted by any student. The proof of the following result uses the principle of mathematical induction and due to space constraints, we only state the result here and provide the details in (Vallam et al. 2014).

Lemma 1. *Given the initial distribution $\overline{\pi}_i^0$ for $\overline{X}_i(t)$ as $\overline{\pi}_i^0(0) = 1$ and $\overline{\pi}_i^0(\overline{x}) = 0 \ \forall \overline{x} \in \{1, 2, \dots, M\}$. The solution to the above differential equations is given by:*

$$\overline{\pi}_i^t(0) = \frac{\mu}{K_i} + \frac{\lambda_i}{K_i} e^{-K_i t}$$

$$\overline{\pi}_i^t(\overline{x}) = \left(-\left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{K_i} - \sum_{\overline{y}=1}^{\overline{x}-1} \frac{\lambda_i^{\overline{x}} t^{\overline{y}} \mu}{\overline{y}! K_i^{\overline{x}-\overline{y}+1}} + \frac{\lambda_i{}^{\overline{x}+1} t^{\overline{x}}}{\overline{x}! K_i} \right) e^{-K_i t} + \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{K_i} \quad \forall \overline{x} : 0 < \overline{x} < M$$

$$\overline{\pi}_i^t(M) = 1 - \sum_{\overline{x}=0}^{M-1} \overline{\pi}_i^t(\overline{x}), \textit{ where } K_i = (\lambda_i + \mu)$$

An additional factor which determines the quantity of rewards being given is that if the instructor may like to *discount* reward per answer if she is coming too often to the forum. So, we introduce $\delta \in (0, 1)$ as the willingness of the instructor to reward the students and $\delta^{h(\mu)}$ gives the *discounting factor* applied by the instructor for *rewardable* answer, where $h(\mu)$ is an increasing function of μ. A reasonable assumption would be $h(\mu) = \log \mu$. The reward $\overline{r}^{l,i}(\overline{x})$ received by a student i of A_l when the instructor visits the forum and finds $\overline{X}_i(t)$ (the lumped CTMC corresponding to student i) in a state $\overline{x}$ is defined as:

$$\overline{r}^{l,i}(\overline{x}) = \begin{cases} \overline{x}\delta^{\log \mu} & if\ \overline{x} \leq m_l, \\ m_l \delta^{\log \mu} & o/w. \end{cases} \tag{2}$$

The net-reward to a student i of A_l in a state $\overline{x}$ will be $\overline{R}^{l,i}(\overline{x}) = \overline{r}^{l,i}(\overline{x}) - \alpha_l \overline{x}$. The expected net-reward at time t to student i of A_l using the lumped-CTMC $\overline{X}_i(t)$ will thus be given by $\overline{R}_t^{l,i} = \sum_{\overline{x} \in \overline{S}_i} \left(\overline{R}^{l,i}(\overline{x})\right) \overline{\pi}_i^t(\overline{x})$ where $\overline{\pi}_i^t(\overline{x})$ is the transient probability being in state $\overline{x}$ at time t. Let $\overline{r}^{I,i}$ denote the reward accrued to the instructor due to answers posted from student i. If the instructor arrives on the forum when $\overline{X}_i(t)$ is in state $\overline{x}$, the reward she receives will be given by: $\overline{r}^{I,i}(\overline{x}) = \overline{x}\delta^{\log \mu}$. Let cost per arrival of the instructor be denoted by β. Then the net-reward to the instructor will be: $\overline{R}^{I,i}(\overline{x}) = \overline{r}^{I,i}(\overline{x}) - \beta$. The expected transient net-reward to the instructor if she arrives on the forum at time t, due to arrival of student i of A_l using the lumped-CTMC $\overline{X}_i(t)$ will thus be given by $\overline{R}_t^{I,i} = \sum_{\overline{x} \in \overline{S}_i} \left(\overline{R}^{I,i}(\overline{x})\right) \overline{\pi}_i^t(\overline{x})$. The expected transient aggregate net-rewards over time T for the student i of A_l and the instructor w.r.t. the CTMC $\overline{X}_i(t)$ will be $\overline{R}_T^{l,i} = \int_{t=0}^{T} \overline{R}_t^{l,i} dt$ and $\overline{R}_T^{I,i} = \int_{t=0}^{T} \overline{R}_t^{I,i} dt$. respectively. Let us define $\overline{R}_T^{I}$ as the total aggregate reward over time T that the instructor receives from all the n lumped-CTMCs $\overline{X}_i(t)$ $i \in \{1, 2, \dots, n\}$. The instructor values each answer on the forum arriving from all students but can unequally value the contributions from different students. We use c_i to give the bias of the instructor towards answers from student i such that $\sum_{i=1}^{n} c_i = 1$ and $0 \leq c_i \leq 1 \ \forall i \in \{1, \dots, n\}$. As all students belonging to the same type are assumed to be similar, so the instructor will value their arrivals equally $c_i = c_j (= c_l) \ \forall i, j \in A_l$. So, we get the total net-reward received by the instructor from the arrival of all the students on the OEF i.e., $\overline{R}_T^{I} = \sum_{1 \leq i \leq n} c_i \overline{R}_T^{I,i}$. We now provide results which connect important computational quantities in the original and the lumped CTMCs.

Lemma 2. *(a)* $\overline{\pi}_i^t(\overline{x}) = \sum_{x \in \overline{S}_i^{\overline{x}}} \pi^t(x)$

Proof. *(Sketch)* (a) The CTMC $X(t)$ being considered has a finite state space S and the arrival rates of the students and the instructor on the CTMC are already known. Due to finiteness, we can assume there exists a finite number $\widehat{q} < \infty$ which bounds the rate entries in the rate matrices Q and $\overline{Q}_i$ ($\forall i \in \{1, 2, \dots, n\}$). The initial state distributions π^0 and $\overline{\pi}_i^0$ for the CTMCs $X(t)$ and $\overline{X}_i(t)$ respectively are defined as $\pi^0(0, 0, \dots, 0) = 1$, $\pi^0(y) = 0 \forall y \in \mathcal{S} \setminus \{(0, 0, \dots, 0)\}$, $\overline{\pi}_i^0(0) = 1$, and $\overline{\pi}_i^0(\overline{x}) = 0 \forall \overline{x} \in \overline{S}_i \setminus \{0\}$. Note that $\overline{\pi}_i^0(\overline{x}) = \sum_{y \in \overline{x}} \pi^0(y) \forall \overline{x} \in \overline{S}_i$. We thus have

$$\overline{\pi}_i^t(\overline{x}) = \sum_{k=0}^{\infty} e^{-\widehat{q}t} \frac{(\widehat{q}t)^k}{k!} \overline{\pi}_i^k(\overline{x}) \quad = \sum_{k=0}^{\infty} e^{-\widehat{q}t} \frac{(\widehat{q}t)^k}{k!} \sum_{y \in \overline{S}_i^{\overline{x}}} \pi^k(y)$$

$$= \sum_{y \in \overline{S}_i^{\overline{x}}} \sum_{k=0}^{\infty} e^{-\widehat{q}t} \frac{(\widehat{q}t)^k}{k!} \pi^k(y) = \sum_{y \in \overline{S}_i^{\overline{x}}} \pi^t(y)$$

The above results follow from invoking results from lumpability of CTMCs (Sumita and Rieders 1989; Buchholz 1994). Please refer to (Vallam et al. 2014) for detailed explanation of the results. $\square$

Lemma 3. *(a)* $R_t^{l,i} = \overline{R}_t^{l,i}$ *(b)* $R_t^I = \sum_{1 \leq i \leq n} c_i \overline{R}_t^{I,i}$

Theorem 2. *The expected transient aggregate net-rewards over time T received by the students and the instructor is the same when calculated using the original CTMC $X(t)$ or the n lumped CTMCs $\overline{X}_i(t)$ $i \in \{1, 2, \dots, n\}$. i.e.,*
(a) $R_T^{l,i} = \overline{R}_T^{l,i}$ *(b)* $R_T^I = \overline{R}_T^I$.

OEF as a Stackelberg Game

We consider a strategic setting where the instructor's goal is to maximize students' participation in the OEF which results in a better understanding of the subject for the students. The students typically will have commitments towards other courses and hence, their objective will be to maximize the rewards they get from answering questions in the OEF while minimizing their cost. We model the strategic interactions in an OEF as a Stackelberg game where the players are n students enrolled for the course and the instructor offering the online course. The instructor acts as a leader who decides her strategy (rate of arrival) first and the students are the followers who, after observing the instructor's strategy, will finalize their own strategies (rates of arrival) in order to maximize their utilities. We assume the strategy space of the players is finite. The key idea behind the formulation of the Stackelberg game is to link the expected net-rewards of the CTMC (when the strategies are known to the model) to the strategic scenario where the players are optimizing their corresponding utility functions. We make this intuition clear in the following subsections by defining the utility functions of the students and instructor and formulating a bi-level optimization problem which yields the optimal instructor and students' strategies. We denote the finite strategy set of the instructor and the students by G^{ins} (of size v) and G^{st} (of size w) respectively. Let $G^{ins} = \{\Delta_a | a \in \{1,..,v\}, \Delta_a \in [0, \widehat{q}]\}$. Let $G^{st} = \{\Lambda_b | b \in \{1,..,w\}, \Lambda_b \in [0, \widehat{q}]\}$. Let $\phi = (\phi_1, \cdots, \phi_v)$ be any arbitrary mixed strategy for the instructor and correspondingly, let $\psi^i = (\psi^i_1, \cdots, \psi^i_w)$ be any arbitrary mixed strategy for student i. For example, the value ϕ_a is the proportion of times that pure strategy Δ_a is used by the instructor while the value ψ^i_b represents the proportion of times in which pure strategy Λ_b is used by student i. Let $\Omega(G^{ins})$ and $\Omega(G^{st})$ be the probability simplices for instructor and students respectively.

Student Optimization Problem

We define expected net-reward matrix $D^{T,l,i}$ for each student i of type l where each entry of the matrix ($D^{T,l,i}_{a,b}$) denotes the expected transient aggregate net-reward received by student i of A_l when she chooses pure strategy Λ_b and instructor chooses the pure strategy Δ_a. Now, suppose the instructor fixes a (pure/mixed) strategy ϕ and the student i fixes a (pure/mixed) strategy ψ^i. The expected transient aggregate utility $U^{l,i}_T$ to a student i of A_l is $U^{l,i}_T = \sum\limits_{a=1}^{v} \sum\limits_{b=1}^{w} D^{T,l,i}_{a,b} \phi_a \psi^i_b$ where $D^{T,l,i}_{a,b} = \overline{R}^{l,i}_T$ (See Thm. 2).

Proposition 1. *Students i, j belonging to the same type A_l receive equal transient aggregate utility if they choose the same policy i.e. if $\psi^i = \psi^j (= \psi^l)$ for students $i, j \in A_l$ then $U^{l,i}_T = U^{l,j}_T$.*

We now formulate the optimization problem for student i of A_l when the instructor has fixed a strategy ϕ.

$$\psi^{i*} = arg \max_{\psi^i} \sum_{a=1}^{v} \sum_{b=1}^{w} D^{T,l,i}_{a,b} \phi_a \psi^i_b, \text{ s.t. } \psi^i \in \Omega(G^{st})$$

Note that ψ^{i*} is the optimal rate of arrival of student i of A_l in response to instructor strategy ϕ. We note that students belonging to the same type can have different optimal strategies in response to the same instructor strategy as the optimization problem can have multiple solutions. Due to symmetry (with respect to cost α_l) among students belonging to the same type, we assume that students belonging to the same type will choose the same optimal strategy i.e. $\psi^{i*} = \psi^{j*} = \psi^{l*} \, \forall i, j \in A_l$. Hence, instead of solving the student optimization problem for each student i, j belonging to a particular type A_l, we can just solve the student optimization problem for a single representative A_l student (now represented as l) and her optimal policy ψ^{l*} would be followed by each student i of A_l. Thus the optimization problem to be solved by each representative student l ($1 \le l \le L$) will be :

$$\psi^{l*} = arg \max_{\psi^l} \sum_{a=1}^{v} \sum_{b=1}^{w} D^{T,l}_{a,b} \phi_a \psi^l_b, \text{ s.t. } \sum_{b=1}^{w} \psi^l_b = 1, \quad \psi^l_b \ge 0$$

where, $D^{T,l}_{a,b} = D^{T,l,i}_{a,b}$ for any $i \in A_l$.

Instructor Optimization Problem

We define aggregate transient utility U^I_T to the instructor when she has fixed her strategy as ϕ and the n students have fixed their policies as ψ^i ($1 \le i \le n$). We define net-reward matrix $B^{T,I,i}$ for the instructor corresponding to each student i where the entry $B^{T,I,i}_{a,b}$ denotes the aggregate utility (over time T) received by the instructor w.r.t. student i's arrivals, if student i chooses Λ_b and instructor the strategy Δ_a. The expected transient aggregate utility of the instructor is given by $U^I_T = \sum_{a=1}^{v} \sum_{b=1}^{w} \sum_{i=1}^{n} B^{T,I,i}_{a,b} \phi_a \psi^i_b$ where, $B^{T,I,i}_{a,b} = c_i \overline{R}^{I,i}_T$ (See Theorem 2).

Proposition 2. *The net-reward matrices $B^{T,I,i}, B^{T,I,j}$ for the instructor w.r.t. students i, j belonging to A_l have the following property: $B^{T,I,i} = B^{T,I,j} (= B^{T,I,l})$.*

The instructor is the leader, so she chooses her policy ϕ first and then each student (follower) observes the strategy chosen by the instructor and then decides the policy ψ^i:

$$\phi^* = \arg\max_{\phi} \sum_{a=1}^{v} \sum_{b=1}^{w} \sum_{i=1}^{n} B^{T,I,i}_{a,b} \phi_a [\psi^{i*}(\phi)]_b \text{ s.t. } \phi \in \Omega(G^{ins})$$

$\psi^{i*}(\phi)$ is an optimal strategy of student i when ϕ is the instructor policy. Let p^l be proportion of type l students in the class. We know that there are n_l students belonging to A_l. We know that $\psi^{i*}(\phi) = \psi^{j*}(\phi) = \psi^{l*}(\phi) \, \forall i, j \in A_l$. Using Proposition 2 and the student optimization problem,

$$\phi^* = \arg\max_{\phi} \sum_{a=1}^{v} \sum_{b=1}^{w} \sum_{l=1}^{L} p^l B^{T,I,l}_{a,b} \phi_a \psi^{l*}_b$$

$$s.t., \; \psi^{l*} = arg \max_{\psi^l} \sum_{a=1}^{v} \sum_{b=1}^{w} D^{T,l}_{a,b} \phi_a \psi^l_b, \phi \in \Omega(G^{ins}), \psi^l \in \Omega(G^{st})$$

MILP Formulation

The mixed integer quadratic program (MIQP) formulated above can be converted to a mixed integer linear program (MILP) by following the well-known approach (See **Proposition 2** in (Paruchuri et al. 2008)) where they solve a Bayesian Stackelberg game by reducing to an MILP.

Numerical Experiments

We solve the MILP using ILOG-CPLEX (ILOG 2014) software and study the changing dynamics of the student-instructor interactions in an online classroom by varying

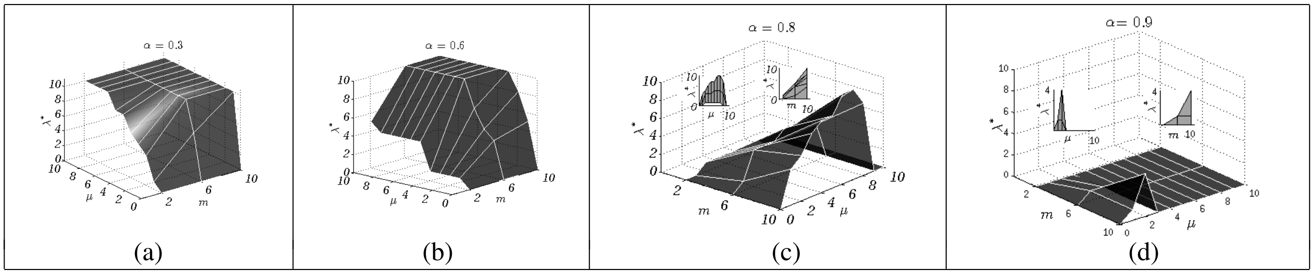


Figure 4: Variation of the optimal participation rates (λ^*) of each student belonging to the different types (characterized by four different α's) with change in the instructor's arrival rate (μ) and the maximum number of rewards (m) given to that type. The $X - Z$ and $Y - Z$ projections of Figures 4c and 4d are given as insets to aid the understanding of the graphs.

the different parameters of the model. We first study the variation in the optimal student participation rate (λ^*) of the heterogeneous student population with μ (instructor arrival rate) and m (budget of instructor) for different student types in Figure 4(a)-(d). Let α denote the cost per arrival to the OEF (and answering an open question if any) of a student. Figure 4(a) represents scenario for an excellent student (i.e., $\alpha = 0.3$). If we fix m, we can observe that λ^* keeps increasing with changing μ. This means that these type of students are quite self motivated and keep posting on the OEF in spite of μ being very high. This is akin to the phenomenon of super-posters (Huang et al. 2014) when there are some students who always aggressively post on the OEF. Figure 4(b) represents scenario for good student (i.e., $\alpha = 0.6$). We can observe a similar behaviour but for low rewards, we can see a dip in the participation rates. Figure 4(c)-(d) denote scenarios for weak students ($\alpha \in \{0.8, 0.9\}$). However, we observe non-monotonic participation patterns for students with high cost per arrival ($\alpha = 0.8, 0.9$) in Figures 4(c) and 4(d) as λ^* initially increases with increasing μ and then, with any further increase in μ, λ^* starts falling. This trend has been noted theoretically (in the homogeneous setting) in literature (Ghosh and Kleinberg 2013).

Now, assume there are only two student types in the class: Type 1 (excellent students) and Type 2 (weak students). We fix these parameters: α_1, α_2, m_1 (budget allocated per question to answers from Type 1 students), m_2 (budget for Type 2 students) where each parameter takes values: $\alpha_1 \in \{0.01, 0.1, 0.2\}, \alpha_2 \in \{0.8, 0.9, 0.99\}, m_1, m_2 \in \{2, 6, 10\}$ resulting in 81 (i.e., $3 \times 3 \times 3 \times 3$) configurations. We first set the instructor bias as a low value ($c_1 = (0.01/n_1)$) for a Type 1 student and as a high value ($c_2 = (0.99/n_2)$) for a Type 2 student where n_1 and n_2 are number of students of Type 1 and 2 respectively. We run the experiment for each configuration separately. As we are dealing with output from multiple experiments, we generate a scatter plot depicting optimal participation rates for a Type 1 student (Figure 5 (a)). We change the instructor behaviour to have high bias towards a Type 1 student and very low bias towards a Type 2 student fixing $c_1 = (0.99/n_1)$ and $c_2 = (0.01/n_2)$ and run the experiments similarly for the 81 parameter configurations as given above. The optimal arrival rates of a Type 1 student in this scenario is given in Figure 5(b). In Figures 5(a)-(b), each point is identified by a number between 1 and 81 and denotes the optimal rate for Type 1 student obtained for the corresponding experimental configuration.

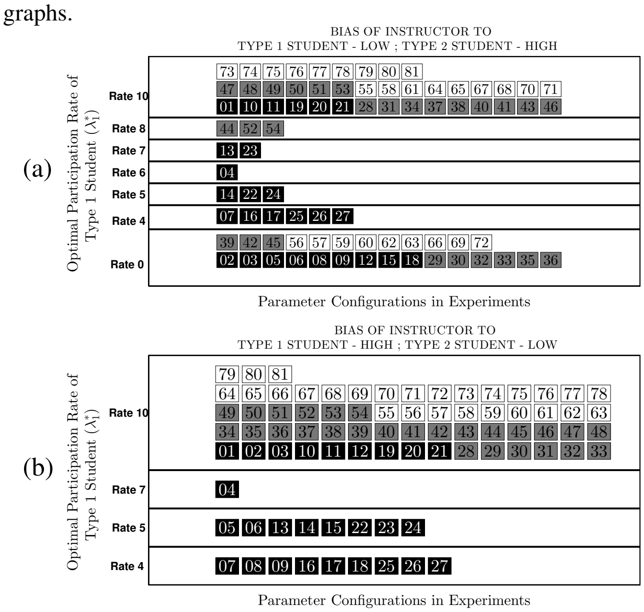


Figure 5: Effect of Instructor bias on the optimal participation rates of a Type 1 student. Each square represents the respective optimal participation rate of a Type 1 student depending on parameter configuration in experiment. Black, Grey and White squares indicate configurations where $m_1 = 2$(low rewards), $m_1 = 6$ (medium rewards) and $m_1 = 10$ (high rewards) respectively.

We observe that, if the instructor's bias towards Student 1 is low then there are configurations when Type 1 student will not participate even for medium and high rewards (for example: configs 45, 56 in Figure 5(a) have optimal rate as 0) whereas if the instructor's bias towards Type 1 students is high, then these students start participating enthusiastically with high rates for medium and high rewards and increase their participation levels even for the lower rewards (for example: configs 45, 56 in Figure 5(a) have optimal rate 10 while config 09 improved from rate 0 in Figure 5(a) to rate 4 in Figure 5(b)). A similar observation can be made for the other type of students. Due to redundancy, we do not show the corresponding graph here. Thus, our model is able to incorporate the effect of instructor bias in deciding the optimal participation level for the different types of students.

As part of our future work, we intend to capture the effect of student's effort for answering a question in the OEF. Further, other ways for modelling OEFs (for example: using Markov decision processes) could be considered. Empirical validations of these results hold promise as well.

Acknowledgements

This research work was partially supported by an IBM Research doctoral fellowship grant to the first author. This work was done when the third author was a project associate in the Game Theory Lab, Department of Computer Science and Automation, Indian Institute of Science, Bengaluru.

References

An, B.; Tambe, M.; Ordóñez, F.; Shieh, E. A.; and Kiekintveld, C. 2011. Refinement of strong stackelberg equilibria in security games. In Burgard, W., and Roth, D., eds., *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, San Francisco, California, USA, August 7-11, 2011*. AAAI Press.

Andresen, M. A. 2009. Asynchronous discussion forums: success factors, outcomes, assessments, and limitations. *Journal of Educational Technology & Society* 12(1).

Brown, M.; Saisubramanian, S.; Varakantham, P.; and Tambe, M. 2014. STREETS: game-theoretic traffic patrolling with exploration and exploitation. In Brodley, C. E., and Stone, P., eds., *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada.*, 2966–2971. AAAI Press.

Buchholz, P. 1994. Exact and ordinary lumpability in finite Markov chains. *J. Appl. Prob.* 31(1):59–75.

Coursera. 2014. Coursera MOOC Platform. Link: https://www.coursera.org/about/community.

Fowler, G. A. 2013. An early report card on massive open online courses. *Wall Street Journal*.

Ghosh, A., and Kleinberg, J. M. 2013. Incentivizing participation in online forums for education. In *ACM Conference on Electronic Commerce*, 525–542.

Huang, J.; Dasgupta, A.; Ghosh, A.; Manning, J.; and Sanders, M. 2014. Superposter behavior in mooc forums. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, 117–126. ACM.

ILOG, I. 2014. *IBM ILOG CPLEX Optimization Studio V12.4*. IBM.

Jain, M.; Tsai, J.; Pita, J.; Kiekintveld, C.; Rathi, S.; Tambe, M.; and Ordóñez, F. 2010. Software assistants for randomized patrol planning for the LAX airport police and the federal air marshal service. *Interfaces* 40(4):267–290.

Mazzolini, M., and Maddison, S. 2003. Sage, guide or ghost? the effect of instructor intervention on student participation in online discussion forums. *Computers & Education* 40(3):237–253.

Paruchuri, P.; Pearce, J. P.; Marecki, J.; Tambe, M.; Ordonez, F.; and Kraus, S. 2008. Playing games for security: an efficient exact algorithm for solving bayesian stackelberg games. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 2*, 895–902.

Sumita, U., and Rieders, M. 1989. Lumpability and time reversibility in the aggregation-disaggregation method for large markov chains. *Stochastic Models* 5(1):63–81.

Vallam, R. D.; Bhatt, P.; Mandal, D.; and Narahari, Y. 2014. A stackelberg game approach for incentivizing participation in online educational forums with heterogeneous student population : Supplementary material. Link:http://lcm.csa.iisc.ernet.in/rohith/AAAI2015/SupplementaryMaterial.pdf.