IMPROVING TEACHER-STUDENT INTERACTIONS IN ONLINE EDUCATIONAL FORUMS USING A MARKOV CHAIN BASED STACKELBERG GAME MODEL [†]

A PREPRINT

Rohith Dwarakanath Vallam [1], Priyanka Bhatt [2], Debmalya Mandal[3], and Y Narahari[4]

[1]IBM India Research Labs, Bengaluru, India
[2]Walmart Labs, Bengaluru, India
[3]Harvard University, Cambridge, USA
[4]Indian Institute of Science, Bengaluru, India
[1]`rohithdv@gmail.com`, [2]`priyankabhatt91@gmail.com`, [3]`dmandal@g.harvard.edu`,
[4]`hari@csa.iisc.ernet.in`

ABSTRACT

With the rapid proliferation 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. Online learning now takes more than one form, including the use of technology to enhance a face-to-face class, a hybrid class that combines both face-to-face meetings and online work, and fully online courses (popularly known as massive open online courses or MOOCs). Further, online classrooms are usually composed of an online education forum (OEF) where students and instructor discuss open-ended questions for gaining better understanding of the subject. However, empirical studies have repeatedly shown that the dropout rates in these online courses are very high partly due to the lack of motivation among the enrolled students. The objective of our work is to come up with an appropriate analytical model for OEFs and use this model to design an effective incentive-based game to enhance student-instructor participation in these OEFs considering for heterogeneity in the skills among the students as well as the limited budget of the instructor.

We first undertake an empirical comparison of student behavior in OEFs associated with a graduate-level course offered in the Indian Institute of Science during two terms. We identify key parameters dictating the dynamics of OEFs like effective incentive design, student heterogeneity, and super-posters phenomenon. Motivated by empirical observations, we propose an analytical model based on continuous time Markov chains (CTMCs) to capture instructor-student interactions in an OEF. Using concepts from lumpability of CTMCs, we compute steady state and transient probabilities along with expected net-rewards for the instructor and the students. We formulate a mixed-integer linear program which views an OEF as a single-leader-multiple-followers Stackelberg game. Through simulations, we observe that students exhibit varied degree of non-monotonicity in their participation (with increasing instructor involvement). We also study the effect of instructor bias and budget on the student participation levels. Our model exhibits the empirically observed super-poster phenomenon under certain parameter configurations and recommends an optimal plan to the instructor for maximizing student participation in OEFs.

---

1   Introduction

With the rapid proliferation 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. Online learning now takes more than one form, including the use of technology to enhance a face-to-face class, a hybrid class that combines both face-to-face meetings and online work (which is typical of courses in many universities today), and online courses (popularly known as massive open online courses or MOOCs). Broadly, any academic course may thus be modelled in three ways: the traditional *face-to-face* model, an *online* model and the *hybrid* model constituting both face-to-face interactions as well as online interactions between students and the instructor. Traditional face-to-face classrooms are identified by an instructor who is conducting the course and students enrolled to the course. Periodic lectures, tests and assignments require the physical presence of the students and the instructor in the classroom at some fixed timings. Hybrid and online classrooms offer a great deal of flexibility in conducting the class which was not possible in the pre-Internet era. One of the forms of online education that has garnered active interest of researchers due to their immense popularity are the massive open online courses (MOOCs) offered by MOOC platforms like Coursera, edX, Udacity, etc. As of January 19, 2017, Coursera [Coursera, 2017] had about 23 million enrollments from students representing 190 countries. Further, Coursera students (as of January 2014) voiced themselves in 590,000 discussion threads in the education forums for a total of 343,014,912 minutes of learning across 571 courses [Butean et al., 2015].

A recent US Department of Education report [Means et al., 2009] made some key observations regarding effectiveness of online learning namely:

- Students taking online courses performed better, on an average, than those taking the same course in a traditional face-to-face environment.

- Hybrid courses having both an online instruction as well as face-to-face elements had a larger advantage relative to purely face-to-face instruction than did purely online instruction.

- The efficacy of online classrooms varies across different courses as well as *heterogeneous skills* of students.

Many hybrid and online courses typically have an online discussion forum (also known as online education forum (OEF)[Ghosh and Kleinberg, 2013]) wherein students with heterogeneous skills can interact with each other as well as the instructor and discuss a variety of questions related to the subject. Such environments offer a great deal of promise for education as well as pose a number of challenges. Surveys of student perceptions about online threaded discussions have shown tremendous positive interest among the students to access these OEFs as an effective way for acquisition of knowledge ([Pendry and Salvatore, 2015],[Martyn, 2003]). Studies in this area have concluded that developing an actively participating learning community along with the instructor participating as an equal member is the vehicle through which online education is best delivered [Palloff and Pratt, 2007]. An important part of such hybrid and online educational environments are OEFs which aim at filling the 'physical' connect missing in these online platforms. Students and the instructor use these forums in order to interact and discuss various topics related to the course and otherwise, which helps in improving the understanding of the students [Andresen, 2009].

It is also noted that participation in OEFs provides opportunities for responsibility and active learning through the expectation of regular participation in online discussions [Hopperton, 1998]. This is not a scenario for a few prepared students to respond to a lecturer while the rest of the class sits back. Participation in the OEF demands that students become actively engaged with the course content and through the interaction with their peers, negotiate the meanings of the content. They construct knowledge through the shared experiences that each participant brings to the collaborative discussions [Markel, 2001]. Thus, the primary objective of online classrooms is to enhance learning by stimulating active discussions on their OEFs and this is determined by the extent of student participations in these OEFs. In this work, we are addressing the problem of improving student participation in the OEFs . In particular, we quantify student participation in OEFs by measuring the number of answers given by enrolled students (with heterogeneous skills) to a number of open-ended questions posted by the instructor on these OEFs. Recent approaches to this problem have focused on an incentive-based approach for improving participation in OEFs [Ghosh and Kleinberg, 2013] and Q&A forums like Yahoo! Answers, Stackoverflow, WikiAnswers, etc ([Ghosh and McAfee, 2012],[Ghosh and Hummel, 2012],[Sakurai et al., 2013]).

In this work, we adopt a game-theoretic approach to address the issue of improving participation levels among students (with heterogeneous skills) in an OEF. We incentivize students on a per question basis to keep up the momentum of participation in the class. Students are required to post answers to the specific open-ended/discussion-style questions that are posted on the OEF. These questions have the characteristic that they do not have a well-defined answer and hence necessitate multiple viewpoints/opinions. Students are thus incentivized to post answers to such questions even if other classmates have already posted their answers resulting in increased student participation and better understanding

of the course content. In practice, these incentives could manifest in the form of book vouchers, food coupons or some extra grade points, as considered appropriate by the instructor.

2 Online Education Forums - Key Issues and State of the Art

Before we get into the details of our approach, we briefly discuss various properties of online classrooms as well as some important issues faced by online education forums (OEFs) in particular. We follow this with a review of some relevant game theoretic literature in this area.

2.1 OEFs, Instructor Participation, and Open-Ended Questions

Numerous studies indicate that OEFs are instrumental in online classrooms for the construction of knowledge and for improving the understanding of the enrolled students. Andresen [Andresen, 2009] and Mazzolini et.al. [Mazzolini and Maddison, 2003] emphasize the critical role of discussion forums and optimal instructor participation in improving the effectiveness of online learning. Breslow et. al [Breslow et al., 2013] observe the data of participation patterns of students in a MOOC offered by edX and their study indicates that 52% of the certificate earners were active on the discussion forums. Williams et. al [Williams et al., 2014] and Ballou [Ballou, 2008] emphasize that open-ended questions help building a rapport among the participants and also improve their understanding. Also, an experimental study by Hull et.al [Hull and Saxon, 2009] suggests that open-ended questions and controlled instructor participation on the OEFs has positive effects on the learning of students.

2.2 Heterogeneous Students

Students participating in MOOCs have been classified in the literature based on various criteria. DeBoer et.al.[DeBoer et al., 2013] identified that students enrolled for MOOCs differ based on geographical locations, educational backgrounds and age groups.

Kizilcec et. al [Kizilcec et al., 2013] observe that there are typically four types of students in online courses based on their level of commitment to the course namely: 'Completing', 'Auditing', 'Disengaging', and 'Sampling'. Wilkowski et. al [Wilkowski et al., 2014] note that there are about four main categories of students namely: 'No-shows', 'Observers', 'Casual Learners', and 'Completers', classified based on their objective of joining the online classroom.
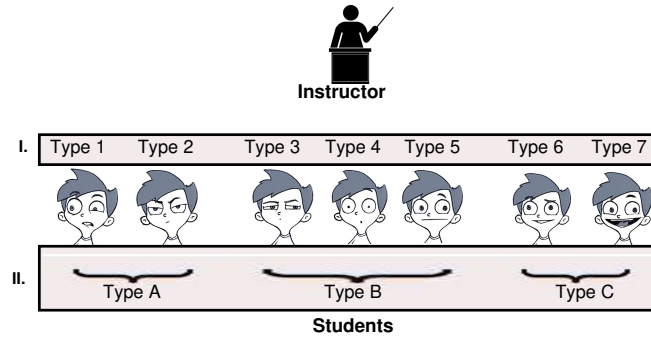

Figure 1: This figure shows a classroom and two sample ways in which the students in a class can be categorized: **I.** All students are different essentially and so each student belongs to a different type. **II.** Sets of students exhibiting properties more similar to each other but different from the rest are put together in one type. (Image Courtesy: BYU Animation)

One of the primary objectives of our work is to model such heterogeneity among the student population by categorizing the students into L types according to some suitable criteria, for example, the expertise of a student in the related areas or the quality of answers given by the students, etc. The classification of students into different types (Figure 1) can be as fine as each student belonging to a different type, or as coarse as all students belonging to a single type. This can depend on the size of the class, for example, in a small class each student can be considered as belonging to a different type whereas in a larger class similar students can be grouped together into one type to reduce the number of types. This grouping is essential in large classrooms (of the scale of online classrooms) as giving personal attention to each student is a humongous task for the instructor and it is easier for the instructor to focus based on grouping of the students.

2.3 Low Student-Retention Rate

Online classrooms witness massive enrollments but their retention rate is very dismal with around 90% students not finishing the course [Fowler, 2013]. Halawa et. al [Halawa et al., 2014], and Cheng et. al [Cheng et al., 2013] claim that a timely intervention by the instructor might help decreasing the dropout rates in the online classrooms and so design predictors and tools which will assist the instructor in identifying the students at a high risk of dropping out.

2.4 Incentives for Participation in OEFs

The instructor of an online course might encounter distinct types of individuals in the OEFs and her objective is to maximize participation from each of them whereas the students' objective is to maximize rewards received (if any) and minimize the cost incurred towards answering questions on the forum. The proposed model thus introduces incentives in the form of some socio-psychological rewards or some physical rewards that are offered to the students on answering a question. The socio-psychological rewards are in the form of satisfaction of the students on getting recognition from the instructor in the classroom and also that the instructor is reading their 'good' answers (we assume that there are teaching assistants for the course who allow only a good quality answer to be posted). On the other hand, physical rewards may be in the form of some book vouchers or food coupons etc. that the instructor can provide to the students. An important point to note is that both types of rewards can be awarded only on a limited scale as the instructor can read or reward only some limited number of answers to a question. This real-world limitation drives us to consider a *budgeted* rewarding scheme which ensures that the instructor has a limited budget B per question, for giving out rewards to the students.

2.5 Review of Relevant Literature

Brafman et.al. [Brafman and Tennenholtz, 1996] analyze partially controlled multi-agent systems and in this context, the authors consider the scenario of a two-agent system consisting of a teacher and a student. The work assumes the teacher is a knowledgeable agent, while the student is an agent that is learning how to behave in its domain. Their goal is to utilize the teacher (which is under our control) to improve the behavior of the student (which is not controlled by us). Our work is similar in philosophy to the above work in that we are proposing an optimal plan to instructor (or teacher) to improve student behaviour in online educational forums. However, we consider solving the problem through designing of optimal incentive scheme in the presence of a budgeted instructor. Also, we customize the behaviour of the instructor taking into account the presence of students with heterogeneous skills.

Fave et.al [Delle Fave et al., 2014] propose a Stackelberg game theoretic model to handle execution uncertainty in the security patrolling domain. They come up with equilibrium strategies for attacker and defenders and perform extensive experiments which implements the proposed model. Our work also models a Stackelberg game but in the domain of education. Our work models an educational forum involving instructor and students in an online course and captures heterogeneity in the skills of the students through a Continuous time Markov Chain model for the online educational forum. Further, the parameters of the CTMC are optimized by viewing the classroom activities as a Stackelberg game model among the instructor and students. Equilibrium strategies are computed which provide an optimal plan to the instructor for maximizing student engagement in an online course.

Ghalme et.al [Ghalme et al., 2018] consider the problem of designing a robust credit score function in the context of online discussion forums. They design coalition resistant credit score functions for online discussion forums. They show that modularity is coalition identifying and provide theoretical guarantees on modularity based credit score function. Carbonara et.al [Carbonara et al., 2015] model the problem of strategic auditing in peer grading modeling the student's choice of effort in response to a grader's audit levels as a Stackelberg game with multiple followers. In this theoretical analysis, a homogeneous student population is considered and provide a polynomial-time approximation scheme (PTAS) in order to compute an approximate solution to the problem of allocating audit levels.

Game theoretic literature that relates to online learning and discussion forums mostly concentrates on designing incentives for Q&A forums like Yahoo! Answers, Stackoverflow, WikiAnswers etc. Ghosh et.al [Ghosh and McAfee, 2012] study about two types of rewards, attention rewards and fixed-total rewards, and how they should be set in order to achieve good quality outcomes from the crowd in an online forum. Sakurai et. al [Sakurai et al., 2013] make the agents indirectly report their efforts and allocate rewards such that agents have to be truthful and have to put in more efforts in order to receive higher rewards. Ghosh et.al [Ghosh and Hummel, 2012] investigate two settings, one in which a contribution's value depends upon an agent's expertise only, and the other in which a contribution's value depends upon both the expertise of the agent and the effort put in by her. They show that optimal outcomes can be implemented in the first case, whereas in the second case, optimal outcomes can be implemented only if the ranking of the contributions is noisy.

Ghosh et.al [Ghosh and Kleinberg, 2013] focus on online forums for education which is closer to our work. They allow the students and the instructor to choose their rates of arrival in order to maximize their utility. Each type of arrival is modeled as a Poisson process and then analysis on the relationship between the students' rate of arrival and the instructor's rate of arrival is done for two types of questions: single-answer type and open-ended type. They are able to show that the students rate of participation varies non-monotonically with the instructor's rate of response to questions, as is generally observed. That is, the students' rate of participation initially increases as the rate of arrival of the instructor increases and then as the instructor's rate increases beyond a threshold, their rates start dropping until they stop participating.

Some key assumptions in Ghosh et.al [Ghosh and Kleinberg, 2013] are homogeneous student population and unlimited number of rewards given to the students, which are very unrealistic in practical settings. Our model generalizes their model in the following aspects: (a) Heterogeneity of student population: we consider a realistic model where students possess heterogeneous skill levels. (b) Budgeted Rewards: the instructor has a limited budget to award students who are participating in the OEF. The analysis in Ghosh and Kleinberg 2013 assumes that all students are similar and this special case can be handled in our model by assuming that all the students belong to the same type.

3 Contributions and Outline

We list the main contributions of this work below:

- We perform empirical analysis of a real-world online education forum by investigating the participation levels of students in the Game Theory course (E1 254) offered in the Indian Institute of Science, Bangalore during Jan-April, 2014. We design incentives to promote better participation in the online Piazza [Piazza, 2017] forum associated with the course. We account for the heterogeneity among the students of the class through various parameters and study the improvement in participation levels in this course compared to the same course offered two years ago where no such incentives for participation were offered.

- We devise a recipe for the instructor in an online course to increase student participation in OEFs. To achieve this goal, we propose a mixed integer linear program (MILP) framework to recommend an optimal plan to the instructor for maximizing student participation in OEFs. We approach this objective through the following intermediate steps:

  (a) Motivated by the empirical study of the online classroom, we first analytically model the activities of an online education forum (OEF) in which there is a course instructor and students with heterogeneous skill levels. In order to promote a higher level of participation by the students in the forum, the instructor periodically posts open-ended (discussion type) questions pertaining to the course in the forum. We propose a continuous time Markov chain (CTMC) to model the non-strategic (i.e. when the rates of arrival of the students and the instructor to the forum are known) student-instructor interactions in an OEF. We model the cost-benefit trade-offs of the students for participating in the forum as well as the limited budget of the instructor by defining corresponding reward functions on the CTMC.

  (b) Using ideas from lumpability of Markov chains, we simplify the CTMC into simpler, student-specific lumped CTMCs. We then compute the steady state and the transient probability distributions of the lumped CTMCs and evaluate the expected net rewards to the instructor and students.

  (c) We then formulate a strategic-OEF setting where both the instructor and the students have a set of possible rates of arrival to choose from and each of them will choose an arrival rate in order to maximize their respective (transient/steady state) utilities. Using a Stackelberg game formulation, we propose a bi-level optimization problem between the instructor and the students in the form of a mixed integer quadratic program (MIQP) which we convert to a mixed integer linear program (MILP) using an approach similar to the analysis done for the Bayesian Stackelberg setting by Parachuri et.al 2008.

  (d) We undertake detailed simulations and develop new insights into the activities of OEFs. We observe that the students exhibit varied degree of non-monotonicity (with increasing instructor rate) in their participation. We also study the effect of instructor bias and budget on the student participation levels. Our model exhibits the empirically observed super-poster phenomenon under certain parameter configurations and recommends an optimal plan to the instructor for maximizing student participation in OEFs.

The rest of the paper is organized as follows. We discuss the empirical analysis of a real-world online education forum in Section 4. We analytically model the activities of an online education forum (OEF) using a CTMC in Section 5. Lumpability of this CTMC is discussed in Section 6. A Stackelberg game formulation is provided in Section 7. Simulation results are discussed in Section 8. We finally conclude in Section 9 and discuss possible avenues of future work in Section 10. The proofs for all results are provided in the appendix.

4   Participation in Online Education Forums - A Case Study

We conducted an experiment using the Piazza [Piazza, 2017] education forum to study the effect of incentives on improving participation in the online classrooms. As part of this experiment, we created user accounts on Piazza for the students, instructors and teaching assistants of the ongoing offline course on Game Theory (E1 254) in the Department of Computer Science and Automation, Indian Institute of Science, in Spring 2014. An explicit incentive system was formulated where participation on forum was rewarded with points. These points could in turn be encashed as an increase of upto 5% in the marks awarded in the offline course of Game Theory. The details of the OEF statistics and the incentives offered are given in Figure 2(a). In order to further incentivize participation, we also organized polls every few weeks on the 2014 OEF to choose the students who were actively posting questions and answers on the forum. The winners of these polls as well as the students who voted for the winners were given some additional reward points. We compared the participation levels we achieved in this forum with data obtained from the OEF for the same course in Spring 2012 at the Indian Institute of Science wherein there was no such incentive designed for improving participation.

OEF Statistics

E1 254	**2012**	**2014**
# Enrolled Students	72	38
Forum Duration	21st Jan to 1st May	1st Mar to 1st May

Incentive System (2014)

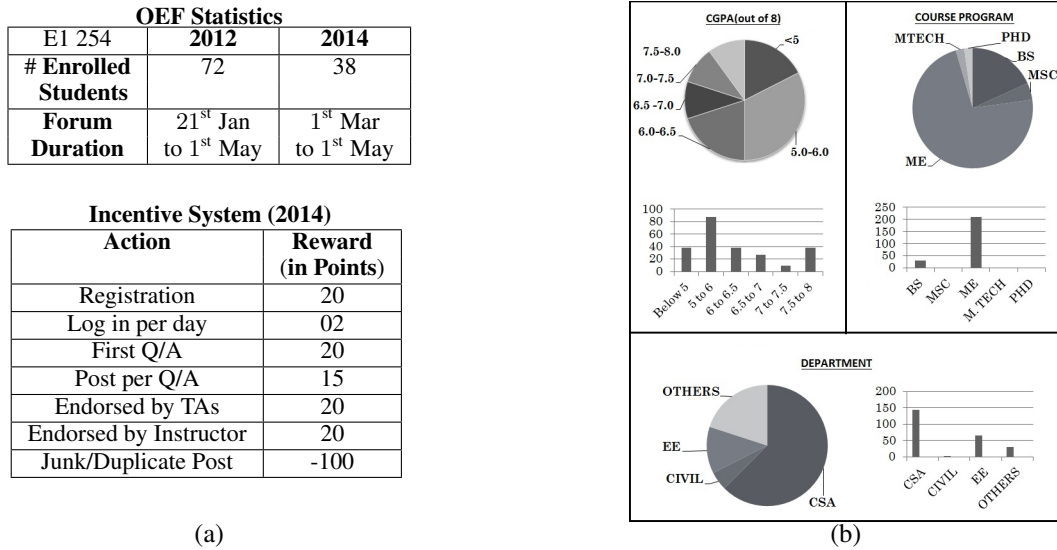| **Action**	**Reward (in Points)**
Registration	20
Log in per day	02
First Q/A	20
Post per Q/A	15
Endorsed by TAs	20
Endorsed by Instructor	20
Junk/Duplicate Post	-100

(a)

(b)

Figure 2: Figures 2(a) to (b) correspond to empirical observations in the Piazza forums of the Game Theory Course for two terms (2012 and 2014) held at the Department of CSA, IISc. Note that the y axis labels in the bar graphs in this figure correspond to "Number of Answers posted on the Piazza forum".

We put forth some of the key observations from our study of the OEFs from these two courses.

- The E1 254 (2014) OEF witnessed high diversity in terms of CGPA, course program and department of the different students that enrolled for the online forum. The different types of students also showed variation in their levels of participation on the forum (See Figure 2(b)).

- The participation levels on the 2014 OEF were higher as compared to the 2012 OEF in terms of the number of questions and answers posted by the enrolled students (See Figure 3(a)). This is despite the fact that the number of enrolled students and the forum duration were lower for the 2014 OEF. Figure 3(b) gives a fine-grained view of the difference in the participation levels in the two OEFs. We thus conclude that incentives that were offered in the 2014 OEF had a major impact on improvement of participation levels in the 2014 class.

- We witnessed different levels of participation from the enrolled students with the total number of posts by a student varying from 0 to mid-30's (See Figure 3(c)). The students also vary in the amount of effort they are willing to put for participation in the online forums. This effort is subjective to a student based on various factors like her capabilities and commitments towards other courses etc. We thus roughly categorize the students as low, medium, high and very high cost (of effort) incurring students depending on their total participation levels through the forum duration (See Figure 3(c)).

- Low cost (of effort) incurring students posted more than 30 posts on the forum and were very active throughout the forum duration irrespective of participation levels of other students and the instructor. This behavior in online forums has also been observed by Huang et. al
  citepHUANG14 and they term such students as "superposters" (marked in a black oval shape in Figure 3(c)).
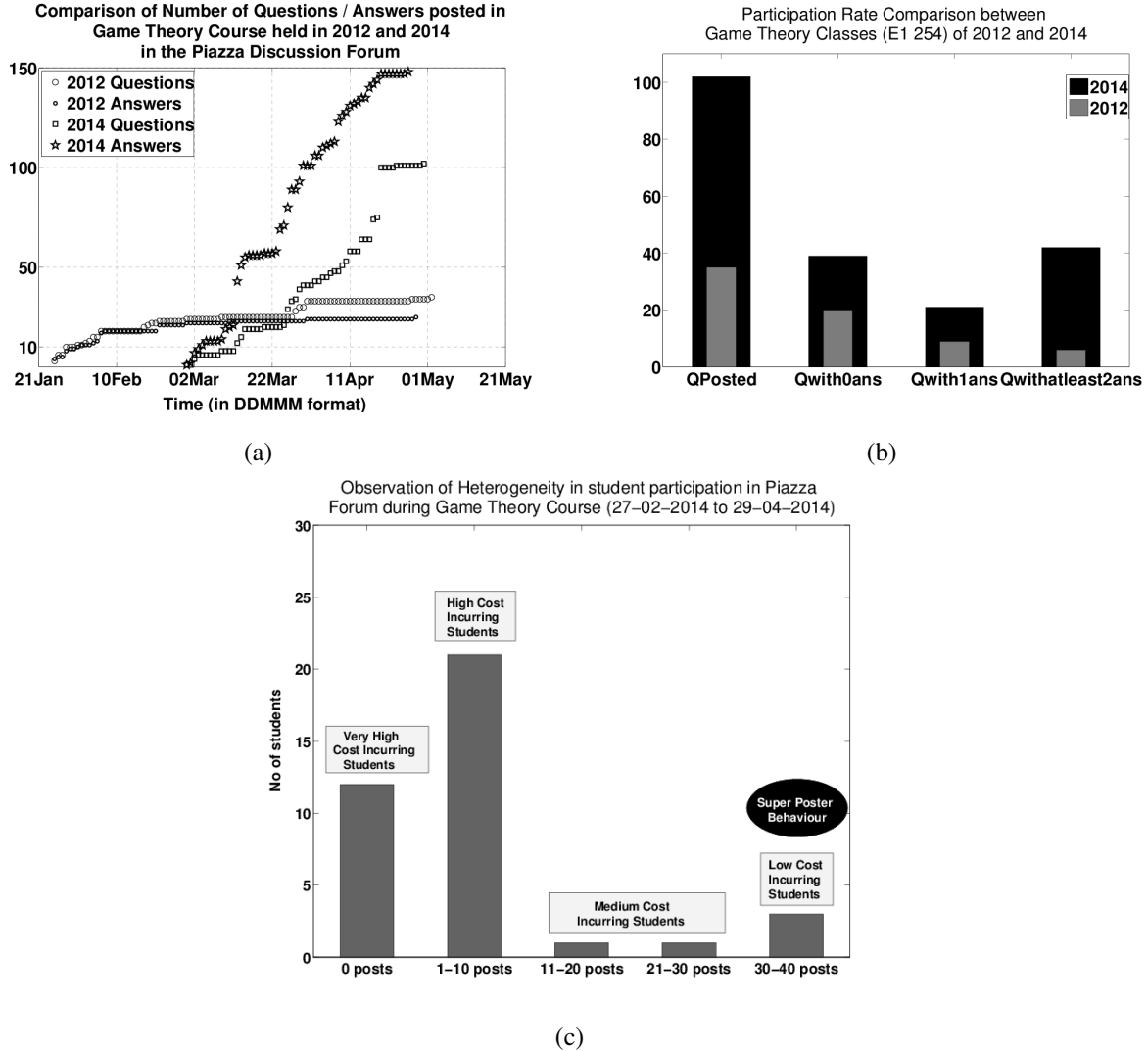
(a)

(b)

(c)

Figure 3: Figures 3(a) to (c) correspond to evaluations of participation levels in the Piazza forums of the Game Theory Course for two terms (2012 and 2014) held at the Department of CSA, IISc.

The incentives that were fixed for the OEF of the E1 254 course (offered in 2014) were decided by taking some basic guidelines as those offered in the Yahoo! Q&A forums
citepJAIN09 and tweaking them a bit to suit our objective. So, the results we have observed are indicative of the general behavior of the students and the positive role that the incentives play in OEFs. The importance of OEFs towards improving the learning of the students and the issues these forums face along with the case study motivate us to do an extensive mathematical modeling of these OEFs. We understand that we might not be able to incorporate all these real life observations in theory and so, in the mathematical modeling part of the report, we shall make suitable assumptions at appropriate junctures.

5 A CTMC Model for OEF (Non-Strategic Setting)

In this section, we first discuss the setting of an online education forum (OEF) and then propose a CTMC to model the arrivals on this forum. We define a suitable reward structure on top of this CTMC and analyze the net-rewards received by the students as well as the instructor. We assume that the instructor and the students participating in the OEF are all non-strategic agents i.e., their rates of participation on the forum are known to the model.

We consider an online classroom setting with n students and an instructor. We begin by proposing an instructor-driven approach to structure the activities in the OEFs (see Figure 4). Henceforth, we will assume the activities of the OEF

7

1: Instructor chooses an arrival rate to the forum and announces this to the class.
2: Students (of different types) observe the instructor announcement and decide their corresponding rate of arrival.
3: Instructor and the student record the next arrival time based on their corresponding chosen arrival rates.
4: **while** Course has not ended **do**
5:     **if** you are the instructor **then**
6:         **if** there is time available for next arrival **then**
7:             Engage in other activities not related to the forum.
8:         **else**
9:             Close current (open-ended) question on the forum.
10:            Reward points to students who have answered this question.
11:            Post the next discussion-style question.
12:            Record the next time for arrival to the forum.
13:        **end if**
14:    **end if**
15:    **if** you are a student **then**
16:        **if** there is time available for next arrival **then**
17:            Engage in any other activities not related to the forum.
18:        **else**
19:            Post a valid answer to the current open question in the OEF.
20:            Record the next time for arrival to the forum.
21:        **end if**
22:    **end if**
23: **end while**

Figure 4: An instructor-driven interaction model in an OEF

follows as outlined in Figure 4. 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. Another modeling assumption is that only the instructor can post questions on to the OEF. Also, once the instructor arrives on the forum, the current question is closed, the students are rewarded based on the number of answers they have given to this question and a new question is posted. Hence, there is a single active question on the forum at any time which remains open until the next visit of the instructor. Closing a question here means not allowing to post further on that question and henceforth, not awarding for answering the closed question. Hence, the students will not find it beneficial to answer the closed questions and thus, it is assumed that they answer only the currently open question. An interesting future direction could be to model the scenario where questions may be posted by both the students and the instructor.

We capture the heterogeneity among students by allowing L types of students in our OEF. Let $Type_l$ be a set containing the students of type l ($\forall l \in \{1, \ldots, L\}$). Also, let n_l be the number of students of type l ($\forall l \in \{1, \ldots, L\}$) in the classroom s.t. $\sum_{1 \leq l \leq L} n_l = n$. On arrival to the forum, a student answers the currently open question (if any) and hence, incurs a cost. Also, as the students belonging to the same type are similar, we assume that the cost of answering a question will be same for all students $i, j \in Type_l$ and this cost will be denoted by α_l. 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* for which the rewards will be given out, for each student $i \in Type_l$ s.t. $B = \sum_{1 \leq l \leq L} n_l m_l$.

5.1 Description of the Model

We now model the activities of the OEF as a continuous time Markov chain (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[4]. We summarize important notations of this section in Table 1. We enumerate the students in the class as $1, 2, \ldots, n$ and define each state $x \in \mathcal{S}$ as: $x = (x_1, \ldots, x_n)$, where each x_i corresponds to the number of answers received from the ith student for the current question. If $X(t)$ is in some state $x : (x_1, \ldots, x_i, \ldots, x_n)$ and student i gives an answer on the forum then $X(t)$ goes into a state $y : (x_1, \ldots, x_i + 1, \ldots, x_n)$, thus capturing each arrival and reflecting the current forum status through its state space. When the instructor comes on the forum, the current question is closed and a new question is started, thus taking the CTMC to state $(0, \ldots, 0)$. Hence, at any point of time, there is a single active question on the forum which when initially posted by the instructor had no answers from any student. Thus, the CTMC will be capturing arrivals of answers for the single question that is active in the forum at a particular instant of time.

---

[4]For readers not familiar with CTMC, q_{ij} i.e. the (i, j)-th entry of the generator matrix Q governs the rate of transition from state i to state j. To be precise, let $P_{ij}(t) = (X(t) = j | X(0) = i)$, then $d/dt(P_{ij}(t))|_{t=0} = q_{ij}$

Symbol	Meaning
m_l	Maximum number of answers considered by instructor per question to $Type_l$ students.
n_l	Number of $Type_l$ students in the course.
$X(t) = (\mathcal{S}, Q)$	The main OEF CTMC with finite state space $\mathcal{S}$ and generator matrix Q (an example is shown in Figure 3(b)).
Q	Generator matrix of $X(t)$.
M	Maximum answers considered from a particular student per question posted on OEF.
μ	Instructor arrival rate to the OEF.
$\hat{q}$	An upper bound on the entries in Q.
Π	Steady state probability vector of $X(t)$.
π^t	Transient probability vector of $X(t)$.
$x = (x_1, \ldots, x_i, \ldots, x_n)$	A state of $X(t)$.
α_l	The cost per arrival of $Type_l$ student.
$r^{l,i}(x)$	Reward received by the student i belonging to $Type_l$ in state x of $X(t)$.
$R^{l,i}(x)$	Net-reward received by a student i belonging to $Type_l$ in a state x of $X(t)$.
c_i	Bias of the instructor towards answers from student i.
$r^{I,i}(x)$	Reward to the instructor if student i gives x_i answers in state $x : (x_1, \ldots, x_i, \ldots, x_n)$ of $X(t)$.
$r^I(x)$	Reward the instructor receives in some state $x : (x_1, \ldots, x_i, \ldots, x_n)$ of $X(t)$.
$\delta^{\log \mu}$	Discounting factor to the instructor.
β	Cost per arrival to the forum for the instructor.
$R^I(x)$	Net-reward received by the instructor on arriving once to the forum when $X(t)$ is in state x.
π^0	Initial distribution vector for the CTMC $X(t)$.
$R_t^{l,i}$	Expected transient net-reward received by a student i of $Type_l$ at time t in $X(t)$.
$\pi^t(x)$	Transient probability of being in state x at time t in $X(t)$.
R_t^I	Expected transient net-reward of the instructor at time t in $X(t)$.
$R_T^{l,i}$	Expected transient aggregate net-reward of a student i of $Type_l$ in $X(t)$.
R_T^I	Expected transient aggregate net-reward of the instructor R_T^I over time T in $X(t)$.
$R_\Pi^{l,i}$	Expected steady state net-reward for the student i of $Type_l$ in $X(t)$.
R_Π^I	Expected steady state net-reward for instructor in $X(t)$.
$q(x, z)$	Entry in Q whose row corresponds to state x and column corresponds to state z.

Table 1: Important notations related to OEF CTMC

The CTMC being considered is actually an infinite state CTMC but since the course will be offered for a limited time and that the instructor and the students are arriving at some reasonable rates allow us to approximate a reasonably large upper bound M on the number of answers that a single student can post on the forum for each question. This transforms our infinite state CTMC to a finite state CTMC where the 'last' state will be $(M, \ldots, M)$, i.e. from this state, the arrival of only the instructor will be captured and the only transition possible from this state will be to the state $(0, \ldots, 0)$. So, the CTMC state space $\mathcal{S}$ will be defined as $\mathcal{S} = \{(x_1, \ldots, x_i, \ldots, x_n) | x_i \in \{0, 1, \ldots, M\}\}$. We depict a sample CTMC in Figure 5 where there are only two students in the OEF.

Let λ_i denote the rate of arrival of student i to the OEF. The instructor closes the active question on each arrival and starts a new one, so, because of the instructor's arrival, each state is transitioning to state $(0, \ldots, 0)$ with rate μ. Note that at any instant of time there can only be a single arrival and then the CTMC will evolve into the next state. Thus, for any two states $x : (x_1, \ldots, x_i, \ldots, x_n)$, $y : (y_1, \ldots, y_i, \ldots, y_n) \in \mathcal{S}$, the entry in the generator matrix Q for the CTMC $X(t)$ will be defined as:

$$Q(x, y) = \begin{cases} \lambda_j & \text{if } \sum_i |x_i - y_i| = 1 \text{ and } \exists j : y_j - x_j = 1 \text{ (Case i)} \\ \mu & \text{if } y = (0, \ldots, 0) \neq x \text{ (Case ii)} \\ \omega & \text{if } x = y \text{ (Case iii)} \\ 0 & o/w \text{ (Case iv)}. \end{cases} \tag{1}$$

$$\text{where } \omega = \sum_{y' \in \mathcal{S} \setminus \{x\}} -Q(x, y').$$
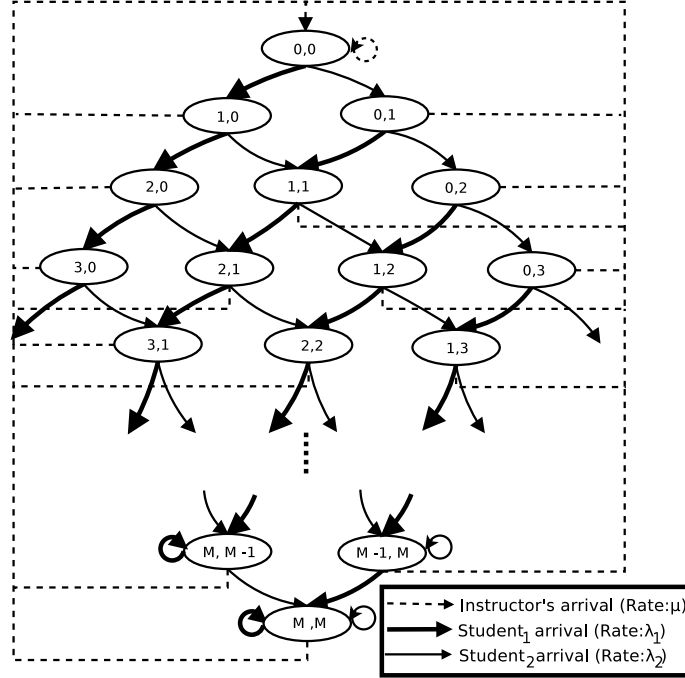
Figure 5: Representative CTMC in our model with 2 students: Each state is capturing the number of answers received from both the students, for a particular question.

Note that due to our finiteness assumption the number of entries in the Q matrix and their values will be finite and thus we can say that each entry in the Q matrix is bounded by a large quantity $\hat{q} < \infty$. Also, let Π and π^t respectively be the steady state (if it exists) and transient state probability vectors of $X(t)$.

5.2 Reward Functions on the CTMC

Each state $x \in X(t)$ indicates the number of arrivals on the OEF by n students. The arrivals of the students and the instructor have some costs and rewards associated with them. In this section, we define a net-reward associated with each state in $\mathcal{S}$ for the students as well as the instructor.

The maximum number of answers for which the rewards will be given out, per question, is fixed for each $Type_l$ student as m_l. So, if a $Type_l$ student answers k times, then the number of answers that are 'rewardable' are $min(k, m_l)$. An additional factor which determines the quantity of rewards being given is that if the instructor might not be willing to give out much reward per answer if she is coming too often and answering the questions herself. So, we introduce $\delta \in (0, 1)$ as the willingness of the instructor to reward the students and $\delta^{h(\mu)}$ gives the actual discounting factor applied by the instructor for 'rewardable' answer, where $h(\mu)$ is an increasing function of μ. For our work, we consider $h(\mu) = \log \mu$. So, if the instructor comes online when the Markov chain is in some state $x = (x_1, \ldots, x_i, \ldots, x_n)$, then the following will be the reward received by the student $i \in Type_l$:

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

The cost per arrival is assumed to be some constant α_l for each answer from a $Type_l$ student. So, if $X(t)$ is in some state $x : (x_1, \ldots, x_i, \ldots, x_n)$, then the cost incurred by a student i belonging to $Type_l$ is $\alpha_l x_i$. Thus the net-reward received by a student i belonging to $Type_l$ in a state x will be given by

$$R^{l,i}(x) = r^{l,i}(x) - \alpha_l x_i. \tag{3}$$

The instructor values each answer on the forum arriving from all students but can unequally value the contributions from different students. Instructor's total reward in a state x from student arrivals will thus be a weighted sum over all students, of the reward she receives on arrival from a single student. 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 $0 \leq c_i \leq 1 \ \forall i \in \{1, \ldots, n\}$. As all students belonging to

the same type are assumed to be similar, so the instructor will value their arrivals equally $c_i = c_j (= c_l) \; \forall i, j \in Type_l$. The instructor intends to maximize the student participation and hence the reward to the instructor if student i gives x_i answers will be given by $r^{I,i}(x) = c_i x_i \delta^{\log \mu}$, where $\delta^{\log \mu}$ is the discounting factor to the instructor. Thus, the reward the instructor receives in some state $x : (x_1, \ldots, x_i, \ldots, x_n)$ will be $r^I(x) = \sum_{i=1}^{n} r^{I,i}(x)$.

Instructor will also incur a cost per arrival to the forum, let us denote this by β. Then the net-reward received by the instructor on arriving once to the forum when $X(t)$ is in state x will be:

$$R^I(x) = \sum_{i=1}^{n} \left(c_i x_i \delta^{\log \mu} \right) - \beta = \sum_{i=1}^{n} c_i (x_i \delta^{\log \mu} - \beta) (\text{as } \sum_{i=1}^{n} c_i = 1)$$

The initial state of the CTMC will be the state $(0, \ldots, 0)$ as the number of answers on the forum from each student will be zero. Thus, the initial distribution π^0 for the CTMC $X(t)$ will be defined as $\pi^0(0, \ldots, 0) = 1$ and $\pi^0(x) = 0 \; \forall x \in \mathcal{S} \setminus \{(0, \ldots, 0)\}$. Now we shall analyze the expected net-rewards of the students and the instructor in $X(t)$. The expected transient net-reward $R_t^{l,i}$ received by a student i of $Type_l$ at time t will be given by :

$$R_t^{l.i} = \sum_{x \in \mathcal{S}} (R^{l,i}(x)) \pi^t(x)$$

where, $R^{l,i}(x)$: Net-reward to student i of $Type_l$ if $X(t)$ is in state x,

$\pi^t(x)$: Transient probability of $X(t)$ being in state x at time t.

We can define the expected transient net-reward of the instructor at time t by:

$$R_t^I = \sum_{x \in \mathcal{S}} (R^I(x)) \pi^t(x)$$

where, $R^I(x)$: Net reward to the instructor if she arrives when

$X(t)$ is in state x

The OEF for an online classroom will be functioning for a certain period of time T (duration of the course) and hence, we need calculate the expected transient aggregate net-rewards over time T and the expected steady-state net-rewards (if the steady state probability vector of the CTMC exists) collected by the students and the instructor. The expected transient aggregate net-reward $R_T^{l,i}$ collected by a student i of $Type_l$ and the expected transient aggregate net-reward collected by the instructor R_T^I over time T will be given by:

$$R_T^{l,i} = \int\limits_{t=0}^{T} R_t^{l,i} \, \mathrm{dt}, \quad R_T^I = \int\limits_{t=0}^{T} R_t^I \, \mathrm{dt}$$

Similarly, the expected steady state net-reward for the students $R_\Pi^{l,i}$ and the instructor R_Π^I will be given by:

$$R_\Pi^{l,i} = \sum_{x \in \mathcal{S}} (R^{l,i}(x)) \Pi(x), \quad R_\Pi^I = \sum_{x \in \mathcal{S}} R^I(x) \Pi(x)$$

where, $\Pi(x)$: Steady-state probability of $X(t)$ being in state x.

As the state space of $X(t)$ is very large ($|\mathcal{S}| = (M + 1)^n$), thus the calculation of the steady-state and transient probabilities of the CTMC $X(t)$ will be very difficult. Thus, computing the expected net-rewards will be very difficult for this complex CTMC and so we shall now use the approach of lumping a CTMC to simplify the analysis of the expected net-rewards for this CTMC.

6   Simplifying the OEF CTMC

In this section, we introduce the concept of lumpability of Markov chains, in the context of discrete time as well as continuous time. We highlight some key results related to lumpability which we will use in reducing the CTMC for the OEF (shown in Figure 3(b)) into simpler student-specific lumped CTMCs. We summarize important notations of this section in Table 2.

Symbol	Meaning
$X(t) = (\mathcal{S}, Q)$	The main OEF CTMC with finite state space $\mathcal{S}$ and generator matrix Q (an example is shown in Figure 3(b)).
$Y_k = (\mathcal{S}, P)$	DTMC with finite state-space $\mathcal{S}$ and probability transition matrix P.
I	Identity matrix.
$P^k(x)$	Probability that the DTMC $(\mathcal{S}, P)$ will be at state x after k steps.
$\overline{S} = \{\overline{S}^j \| j \in \{1, \ldots, w\}\}$	Partition of the state space $\mathcal{S}$ i.e. $\cup_{j=1}^{w} \overline{S}^j = \overline{S}$, $\overline{S}^i \neq \emptyset$, and $\overline{S}^i \cap \overline{S}^j = \emptyset$ for $i \neq j$.
$\overline{Y}_k = (\overline{S}, \overline{\pi})$	DTMC obtained after applying lumping to DTMC $Y_k = (\mathcal{S}, P)$.
$\overline{S}_i$	Partition on the state space $\mathcal{S}$ of $X(t)$ w.r.t. a student i in the OEF i.e., $\overline{S}_i = \{\overline{S}_i^a \| a \in \{0, 1, \ldots, M\}\}$.
$\overline{S}_i^a$	This is the set $\{(x_i, x_{-i}) \in \mathcal{S} \| x_i = a\}$.
$\overline{X}_i(t) = (\overline{S}_i, \overline{Q}_i)$	The quotient (lumped) Markov chain we get on lumping the CTMC $X(t)$ w.r.t. partition $\overline{S}_i$ ($i \in \{1, 2, \ldots, n\}$) (an example is shown in Figure 7).
$\overline{\Pi}_i$	Steady state probability vector for the lumped CTMC $\overline{X}_i(t)$.
$\overline{\pi}_i^t$	Transient state probability vector for the lumped CTMC $\overline{X}_i(t)$.
$\overline{r}^{l,i}(\overline{x})$	The reward received by a student i of $Type_l$ in a state $\overline{x}$ of the $\overline{X}_i(t)$ (the lumped CTMC corresponding to student i).
$\overline{R}^{l,i}(\overline{x})$	Net-reward to a student i of $Type_l$ in a state $\overline{x}$ for the lumped CTMC $\overline{X}_i(t)$.
$\overline{R}_t^{l,i}$	Expected net-reward at time t to student i of $Type_l$ in the lumped-CTMC $\overline{X}_i(t)$.
$\overline{\pi}_i^t(\overline{x})$	Transient probability of being in state $\overline{x}$ at time t in $\overline{X}_i(t)$.
$\overline{R}_t^{I,i}$	Expected transient net-reward to the instructor if she arrives on the forum at time t, from the arrival of student i of $Type_l$ in the lumped-CTMC $\overline{X}_i(t)$.
$\overline{R}_T^{l,i}$	Expected transient aggregate net-rewards over time T for the student i of $Type_l$ w.r.t. the CTMC $\overline{X}_i(t)$.
$\overline{R}_T^{I,i}$	Expected transient aggregate net-rewards over time T for the instructor w.r.t. the CTMC $\overline{X}_i(t)$.
$\overline{R}_{\overline{\Pi}}^{l,i}$	Expected steady state net-rewards over time T for the student i of $Type_l$ w.r.t. the CTMC $\overline{X}_i(t)$.
$\overline{R}_{\overline{\Pi}}^{I,i}$	Expected steady state net-rewards over time T for the instructor w.r.t. the CTMC $\overline{X}_i(t)$.
$\overline{R}_T^I$	Total transient aggregate net-reward received by the instructor from the arrival of all the students on the OEF.
$\overline{R}_{\Pi}^I$	Total steady state net-reward received by the instructor from the arrival of all the students on the OEF.
$\overline{\pi}_i^0$	Initial state distribution for the CTMC $\overline{X}_i(t)$.
$\overline{\Pi}_i(\overline{x})$	Probability that the CTMC $\overline{X}_i(t)$ (corresponding to student i) is in state $\overline{x}$, when $\overline{X}_i(t)$ has reached a steady state.
$\overline{\pi}_i^t(\overline{x})$	Transient state probability of $\overline{X}_i(t)$ being in the state $\overline{x}$ at a time instant t.

Table 2: Important notations related to lumped CTMC

6.1 Lumpability of Markov Chains - Important Results from Literature

6.1.1 Lumpability of DTMC

Consider a DTMC Y_k with finite state-space $\mathcal{S}$ and probability transition matrix P. Let $\overline{S} = \{\overline{S}^j \| j \in \{1, \ldots, w\}\}$ be a partition of the state space $\mathcal{S}$ i.e. $\cup_{j=1}^{w} \overline{S}^j = \overline{S}$, $\overline{S}^i \neq \emptyset$, and $\overline{S}^i \cap \overline{S}^j = \emptyset$ for $i \neq j$. Let us define the stochastic process Z_k as follows

$$Z_k = j \quad \text{if } Y_k \in \overline{S}^j \quad \text{for } j \in \{1, \ldots, w\}$$

Then Z_k is called the lumped stochastic process of Y_k with respect to the partition $\overline{S}$.

Definition 1. *[Sumita and Rieders, 1989] A finite DTMC Y_k on $\mathcal{S}$ is said to be lumpable with respect to the partition $\overline{S}$ if the lumped process Z_k is also a DTMC for any initial probability vector.*

It may be difficult to check Definition 1 for all initial probability vectors. Kemeny et.al [Kemeny and Snell, 1960] provide the following necessary and sufficient condition for lumpability which is easy to verify. For notational ease, we use $p(x, A) = \sum_{z \in A} p(x, z)$ where $A \subseteq \overline{S}$.

Theorem 1. *[Kemeny and Snell, 1960] (i) A finite DTMC $Y_k = (\mathcal{S}, P)$ is lumpable with respect to the partition $\overline{S}$ iff for any two blocks $\overline{S}^a, \overline{S}^b \in \overline{S}$ and for all $x, y \in \overline{S}^a$ we have $p(x, \overline{S}^b) = p(y, \overline{S}^b)$. (ii) In such a case the lumped Markov chain $\overline{Y}_k = (\overline{S}, \overline{P})$ is determined by $\overline{P}(\overline{S}^a, \overline{S}^b) = p(x, \overline{S}^b)$ for some $x \in \overline{S}^a$.*

6.1.2 Lumpability of CTMC

Consider a finite CTMC $X(t)$ with state space $\mathcal{S}$ and generator matrix Q. Analogous to DTMC, one can construct the lumped process of $X(t)$ with respect to a partition $\overline{S}$ of $\mathcal{S}$ and define the lumpability of $X(t)$ with respect to $\overline{S}$. However, the lumpability of CTMC can be characterized in terms of a related DTMC as follows.

Theorem 2. *([Sumita and Rieders, 1989],[Buchholz, 1994]) Consider a CTMC $X(t)= (\mathcal{S}, Q)$. If the transition rates $Q(x, y)$ are bounded by some positive constant $\widehat{q} < \infty$, then there exists a DTMC Y_k with transition matrix $P = Q/\widehat{q}+I$ such that*

- *the distribution vector $\pi(t)$ of the CTMC at time t starting with $\pi(0)$ at time 0 is given by*

$$\pi(t) = \sum_{k=0}^{\infty} \exp(-\widehat{q}t) \left[(\widehat{q}t)^k / k! \right] \pi^k \tag{4}$$

  where π^k is the distribution vector of Y_k after k jumps.

- *the stationary probability vectors of $X(t)$ and Y_k are identical.*

- *$X(t)$ is lumpable with respect to $\overline{S}$ iff Y_k is lumpable with respect to $\overline{S}$.*

Equation 4 states that the probability of being in a state $x \in \mathcal{S}$ at time t for the CTMC $X(t)$ is equal to the sum over all possible number of jumps k of the probability that exactly k jumps occur in t time [8] and that after k jumps the underlying DTMC Y_k will be in state x. For ease of notation, we use $q(x, A) = \sum_{z \in A} q(x, z)$ where $A \subseteq \overline{S}$. An immediate corollary of Theorem 2 is the following :

Corollary 1. *([Sumita and Rieders, 1989],[Derisavi et al., 2003],[Buchholz, 1994]) (i) A CTMC $X(t)=(\mathcal{S}, Q)$ is lumpable w.r.t. a partition $\overline{S}$ of $\mathcal{S}$ iff for any two blocks $\overline{S}^a, \overline{S}^b \in \overline{S}$ and for every $x, y \in \overline{S}^a$ we have $q(x, \overline{S}^b) = q(y, \overline{S}^b)$. (ii) In such a case, the lumped Markov chain $\overline{X}(t) = (\overline{S}, \overline{Q})$ is determined by $\overline{Q}(\overline{S}^a, \overline{S}^b) = q(x, \overline{S}^b)$ for some $x \in \overline{S}^a$.*

6.1.3 Transient and Steady-State Probabilities of the Lumped Process

For a DTMC (S, P), let us define $P^k(x)$ [1] to be the probability that the Markov chain will be at state x after k steps. The following theorem relates the transient probability of the lumped Markov process with the transient probability of the original Markov chain.

Theorem 3. *([Sumita and Rieders, 1989], (Theorem 5 in [Buchholz, 1994])) Let the finite state DTMC $Y_k = (S, \pi)$ be lumpable w.r.t. the partition $\overline{S}$ of S and let $\overline{Y}_k = (\overline{S}, \overline{\pi})$ be the lumped DTMC. Then $\overline{\pi}^k(\overline{S}^a) = \sum_{y \in \overline{S}^a} \pi^k(y)$ $\forall \overline{S}^a \in \overline{S}$, for $\overline{\pi}^0(\overline{S}^a) = \sum_{y \in \overline{S}^a} \pi^0(y)$ $\forall \overline{S}^a \in \overline{S}$.*

Therefore, the probability that the lumped DTMC is in a block after k steps is equal to the sum of k-step transition probabilities of the states belonging to the same block in the original DTMC. The next theorem provides a sufficient condition for the lumped CTMC to have a stationary distribution and connects this distribution with the stationary distribution of the original CTMC.

Theorem 4. *([Tian and Kannan, 2006], §2.12) Let X(t) be an aperiodic, irreducible CTMC with stationary distribution Π. If it is lumpable w.r.t. a partition of the state space, then the lumped chain also has a stationary distribution $\overline{\Pi}$ whose components can be obtained from Π by adding corresponding components in the same block of the partition.*

6.2 Lumpability of the OEF CTMC

Recall the CTMC $X(t) = (\mathcal{S}, Q)$ where $\mathcal{S} = \{(x_1, \ldots, x_i, \ldots, x_n) | x_i \in \{0, 1, \ldots, M\}\}$ and Q is defined as in Equation 1. This CTMC represents the OEF but is very complex to analyze because of its large state space $(|\mathcal{S}| = (M + 1)^n)$. So, we exploit the basic structure of the problem and come up with lumped-CTMCs with smaller

---

[8] $exp(-\hat{q}t)((\hat{q}t)^k/k!)$ is the probability that a Poisson process with rate $\hat{q}$ makes exactly k jumps in time t

[1] We omit the initial distribution $P^0(\cdot)$ in $P^k(\cdot)$, since this will be clear from the context.
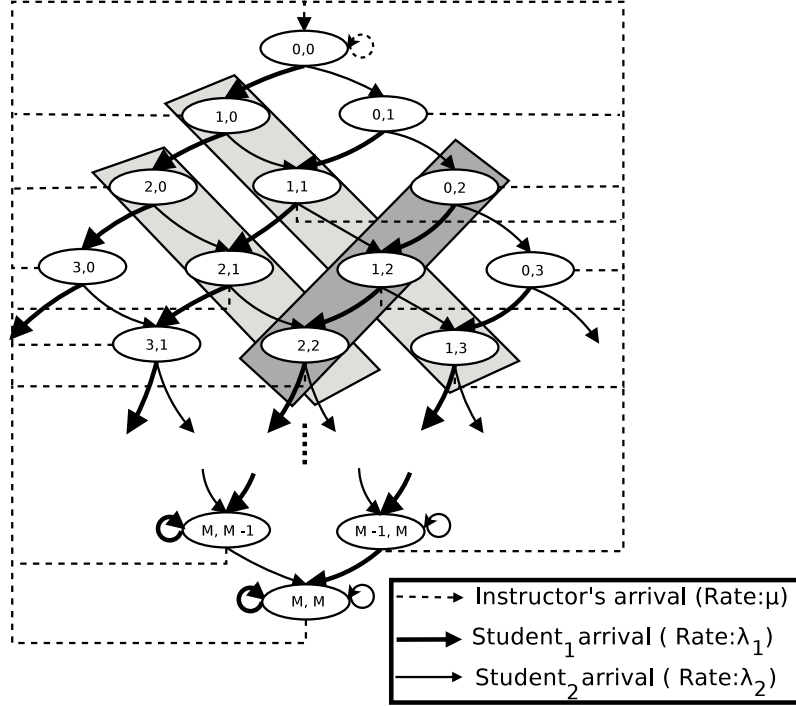
Figure 6: Light-gray and dark-gray quadrilaterals in the figure depict two partitions of the state space that can be defined on the example CTMC given in Figure 5. Each light-gray quadrilateral contains the states where the number of arrivals of student 1 are same. Each dark-gray quadrilateral contains the states where the number of arrivals of student 2 are equal.

state-space which will be easier to analyze. An important characteristic to note about our problem 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 strategy and the instructor's strategy. This is possible because of the property of open-ended questions that they have more than one possible answer 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 complex CTMC, which keeps track of arrivals of all students (of all types), we show that we can analyze n independent player-specific CTMCs (with $M+1$ states each) so that each of the player-specific CTMC keeps track of arrivals from only a particular student. This is possible by applying the lumping process on the original-CTMC which we describe next.

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 $\overline{S}_i^a = \{(x_i, x_{-i}) \in \mathcal{S} | x_i = a\}$.

We lump together all the states in one light gray quadrilateral (having same number of answers from Student 1 in Figure 6 as one state of the lumped chain in Figure 7 (a). For example, the state 1 in Figure 7 (a) basically denotes the set of states $\{(1, 0), (1, 1), \ldots, (1, M)\}$ from Figure 6 and state 2 in Figure 7 (a) denotes the set of states $\{(2, 0), (2, 1), \ldots, (2, M)\}$ from Figure 6, and so on. Similarly, all the states in the dark gray quadrilateral (having same number of answers from Student 2) are represented as one state of the lumped chain in Figure 7 (b). We now present a result which states that, given student i, $X(t)$ is lumpable w.r.t. partition $\overline{S}_i$. Detailed proofs of the results presented in this section are provided in Appendix A.

Theorem 5. [3] *(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 :*
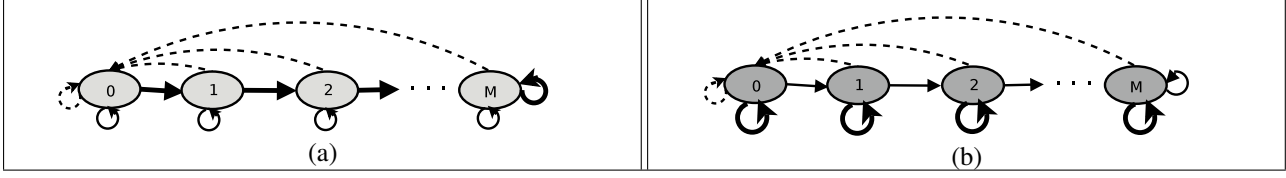
Figure 7: **a)** Lumped-Student$_1$ CTMC : Each light-gray state indicates the aggregation of all the states in the Figure 6 CTMC which are enclosed by the light-gray quadrilateral. **b)** Lumped-Student$_2$ CTMC : Each dark-gray state indicates the aggregation of all the states in the Figure 6 CTMC which are enclosed by the dark-gray quadrilateral.

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

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

So, from Theorem 5, we now have n lumped-CTMCs $\overline{X}_i(t) = (\overline{S}_i, \overline{Q}_i)$, $1 \leq i \leq n$ (See Figure 7) 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. Now, using Theorem 5 (ii) the generator matrix $\overline{Q}_i$ for $\overline{X}_i(t)$ can be re-written as :

$$\overline{Q}_i(\overline{x}, \overline{y}) = \begin{cases} \lambda_i & \overline{y} = \overline{x} + 1, \\ \mu & \overline{y} = 0 \neq \overline{x}, \\ - \sum_{\overline{z} \in \overline{S}_i \setminus \{\overline{x}\}} \overline{Q}_i(\overline{x}, \overline{z}) & \overline{y} = \overline{x}, \\ 0 & o/w. \end{cases} \tag{5}$$

The initial state distribution $\overline{\pi}_i^0$ for the CTMC $\overline{X}_i(t)$ is defined as $\overline{\pi}_i^0(0) = 1$ and $\overline{\pi}_i^0(\overline{x}) = 0 \forall \overline{x} \in \overline{S}_i \setminus \{0\}$ as, initially, no answer would be posted on the forum by any student. Also, let $\overline{\Pi}_i$ and $\overline{\pi}_i^t$ respectively denote the steady state probability vector (if it exists) and the transient state probability vector for the lumped CTMC $\overline{X}_i(t)$.

We shall now define the notions of rewards, net-rewards, and expected net-rewards corresponding to the ith lumped-CTMC, for both the student i and the instructor. The definitions will be very similar to the ones discussed for the original-CTMC (as given in Section 5.2). The reward $\overline{r}^{l,i}(\overline{x})$ received by a student i of $Type_l$ when the instructor visits the forum and finds $\overline{X}_i(t)$ (the lumped CTMC corresponding to student i) in a state $\overline{x}$ will now be defined as:

$$\overline{r}^{l,i}(\overline{x}) = \begin{cases} \overline{x} \delta^{\log \mu} & if \; \overline{x} \leq m_l, \\ m_l \delta^{\log \mu} & o/w. \end{cases} \tag{6}$$

As discussed before (in Section 5.2), a $Type_l$ student will also incur a cost of α_l per arrival therefore the net-reward to a student i of $Type_l$ in a state $\overline{x}$ will be:

$$\overline{R}^{l,i}(\overline{x}) = \overline{r}^{l,i}(\overline{x}) - \alpha_l \overline{x} \tag{7}$$

---

[3]In fact a general form of Theorem 5 is true . Consider a group of students G and define $\bar{S}_G = \{\bar{S}_G^a : a \in \{0, 1, \dots, M\}^G\}$, where $\bar{S}_G^a = \{(x_G, x_{-G}) : x_G = a\}, \forall a \in \{0, 1, \dots, M\}^G$ i.e. we lump together all the states where the entries for the students in group G is fixed at a. Then we can prove that the CTMC $X(t) = (S, Q)$ is lumpable w.r.t. the partition $\bar{S}_G$. Moreover, the lumped Markov chain $\bar{X}_t = (\bar{S}_G, \bar{Q}_G)$ is given as $\bar{Q}_G(\bar{S}_G^a, \bar{S}_G^b) = Q(a, b)$ where $Q(a, b)$ is defined in Equation 1. We omit the proof of this general statement since the proof is quite similar to the proof of Theorem 5 and we will not be using the general lumpability result in our paper.

The expected net-reward at time t to student i of $Type_l$ 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, $\overline{R}^{l,i}(\overline{x})$: Net-reward to student i of $Type_l$ in state $\overline{x}$ of $\overline{X}_i(t)$,

$\overline{\pi}_i^t(\overline{x})$: Transient probability of $\overline{X}_i(t)$ being in state $\overline{x}$ at time t.

As each state of the lumped-CTMC $\overline{X}_i(t)$ captures the number of arrivals from student i, thus we use $\overline{X}_i(t)$ to calculate the reward $\overline{r}^{I,i}$ collected by the instructor by arrivals from student i. As only one student's arrivals are being tracked in $\overline{X}_i(t)$, 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: $\overline{r}^{I,i}(\overline{x}) = \overline{x}\delta^{\log \mu}$ (Refer Section 5.2 for details.). Also, if we take into account the cost per arrival of the instructor 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, from the arrival of student i of $Type_l$ using the lumped-CTMC $\overline{X}_i(t)$ will thus be given by:

$$\overline{R}_t^{I,i} = \sum_{\overline{x} \in \overline{S}_i} \left(\overline{R}^{I,i}(\overline{x}) \right) \overline{\pi}_i^t(\overline{x})$$

where, $\overline{R}^{I,i}(\overline{x})$: Net-reward to the instructor in state $\overline{x}$ of $\overline{X}_i(t)$

from student i's participation,

We shall now define the expected transient aggregate net-rewards over time T and also the expected steady state net-rewards (if steady state probability vector of the lumped CTMC $\overline{X}_i(t)$ exists), to the student i and the instructor w.r.t. the CTMC $\overline{X}_i(t)$.

The expected transient aggregate net-rewards over time T for the student i of $Type_l$ and the instructor w.r.t. the CTMC $\overline{X}_i(t)$ will be denoted by $\overline{R}_T^{l,i}$ and $\overline{R}_T^{I,i}$ respectively. These are the rewards that the student i and the instructor accumulate over time T from the arrivals captured using the lumped-CTMC $\overline{X}_i(t)$ and will be defined as:

$$\overline{R}_T^{l,i} = \int_{t=0}^{T} \overline{R}_t^{l,i} dt, \quad \overline{R}_T^{I,i} = \int_{t=0}^{T} \overline{R}_t^{I,i} dt.$$

The expected steady state net-rewards for the student i of $Type_l$ and the instructor w.r.t. CTMC $\overline{X}_i(t)$ will be the expected net-rewards received by the student i and the instructor when $\overline{X}_i(t)$ is in the steady state and are denoted by $\overline{R}_{\overline{\Pi}}^{l,i}$ and $\overline{R}_{\overline{\Pi}}^{I,i}$ respectively. These will be defined as:

$$\overline{R}_{\overline{\Pi}}^{l,i} = \sum_{\overline{x} \in \overline{S}_i} \overline{R}^{l,i}(\overline{x})\overline{\Pi}(\overline{x}), \quad \overline{R}_{\overline{\Pi}}^{I,i} = \sum_{\overline{x} \in \overline{S}_i} \overline{R}^{I,i}(\overline{x})\overline{\Pi}(\overline{x})$$

where, $\overline{\Pi}(\overline{x})$ =Steady-state probability of $\overline{X}_i(t)$ being in state $\overline{x}$.

As each lumped-CTMC $\overline{X}_i(t)$ captures the arrival of only student i on the OEF, thus the net-rewards received by the instructor in $\overline{X}_i(t)$ is from student i's arrivals only. So lets define $\overline{R}_T^I$, $\overline{R}_{\Pi}^I$ as the total aggregate reward over time T and the total steady state reward that the instructor receives from all the n lumped-CTMCs $\overline{X}_i(t)$ $i \in \{1, 2, \ldots, n\}$. The instructor has a bias c_i for arrivals from student i (as explained in Section 5.2). So, we take a weighted sum over the expected net-rewards received by the instructor in all the n lumped-CTMCs $\overline{X}_i(t)$ $i \in \{1, 2, \ldots, n\}$ to get the total net-reward received by the instructor from the arrival of all the students on the OEF.

$$\overline{R}_T^I = \sum_{1 \le i \le n} c_i \overline{R}_T^{I,i}, \quad \overline{R}_{\Pi}^I = \sum_{1 \le i \le n} c_i \overline{R}_{\Pi}^{I,i}.$$

The next lemma proves that the transient probability $\overline{\pi}_i^t(\overline{x})$ of $\overline{X}_i(t)$ being in state $\overline{x}$ at time t is obtained by summing the transient probabilities $\pi^t(x)$ of $X(t)$ being in state x at time t over all x belonging to the block $\overline{S}_i^{\overline{x}}$.

Lemma 1. $\overline{\pi}_i^t(\overline{x}) = \sum_{x \in \overline{S}_i^{\overline{x}}} \pi^t(x)$

The following lemma proves that the expected transient net-reward to the student i of $Type_l$ is same when calculated using the CTMCs $X(t)$ and $\overline{X}_i(t)$. Thus instead of calculating the net-reward to a student i using $X(t)$, we can calculate the net-reward to the student i using the ith lumped-CTMC $\overline{X}_i(t)$.

Lemma 2. $R_t^{l,i} = \overline{R}_t^{l,i}$

The following lemma proves that the expected transient net-reward to the instructor is the same when calculated using the CTMC $X(t)$ and when calculated as a weighted sum of the instructor's expected transient net-rewards w.r.t. CTMCs $\overline{X}_i(t)$ for student $i : 1 \leq i \leq n$. Therefore we can utilize the lumped CTMCs $\overline{X}_i(t)$ instead of the original CTMC $X(t)$ to calculate the instructor's expected transient net-reward.

Lemma 3. $R_t^I = \sum\limits_{1 \leq i \leq n} c_i \overline{R}_t^{I,i}$

Next we prove that the original CTMC $X(t)$ and the lumped CTMCs $\overline{X}_i(t)$ have steady state vectors and provide a relationship between them.

Lemma 4. *The CTMC $X(t) = (\mathcal{S}, Q)$ is irreducible and positive recurrent and a steady state vector Π exists for this CTMC.*

Lemma 5. *The lumped-CTMC $\overline{X}_i(t) = (\overline{S}_i, \overline{Q}_i)$ is irreducible and positive recurrent and a steady state vector $\overline{\Pi}_i$ exists for this CTMC. Also each component of $\overline{\Pi}_i$ will be obtained as: $\overline{\Pi}_i(\overline{x}) = \sum\limits_{x \in \overline{S}_i^{\overline{x}}} \Pi(x)$.*

Theorem 6. *The expected transient aggregate net-rewards over time T and the expected steady-state net-rewards received by the students and the instructor is the same when calculated using the original CTMC $X(t)$ or the n lumped CTMCs $\overline{X}_i(t)$ $i \in \{1, 2, \ldots, n\}$.*

- *Expected transient aggregate net-rewards over time T.*
  a) $R_T^{l,i} = \overline{R}_T^{l,i}$ b) $R_T^I = \overline{R}_T^I$.

- *Expected steady-state net-rewards.*
  a) $R_\Pi^{l,i} = \overline{R}_\Pi^{l,i}$ b) $R_\Pi^I = \overline{R}_\Pi^I$.

Thus, the expected transient net-rewards to the students and the instructor at time t can be calculated using the n lumped CTMCs $\overline{X}_i(t)$ ($1 \leq i \leq n$) instead of using the original complex CTMC $X(t)$. So, for our analysis, we shall now focus on the lumped-CTMCs for each student instead of the original-CTMC.

6.3 Analysis of the Lumped CTMCs

Recall that each lumped-CTMC $\overline{X}_i(t) = (\overline{S}_i, \overline{Q}_i)$ for student i of $Type_l$ will have the following entities associated with it: max-arrival M, max-reward m_l, course duration time T, rate of arrival λ_i for the student i in the OEF and the rate of instructor arrival μ fixed by the instructor. Also recall that the state space $\overline{S}_i$ of $\overline{X}_i(t)$ is given by $\overline{S}_i = \{0, 1, \ldots, M\}$ and the generator matrix $\overline{Q}_i$ is defined as given in Equation 5. Each lumped-CTMC will thus be similar in structure to the CTMC given in Figure 7 (a).

We give the matrix representation of the generator matrix $\overline{Q}_i$ below for better clarity. Each entry of this matrix is bounded by a large value $\widehat{q} < \infty$(Discussed in Section 5.1). Also recall that the initial state distribution $\overline{\pi}_i^0$ for the CTMC $\overline{X}_i(t)$ is defined as $\overline{\pi}_i^0(0) = 1$ and $\overline{\pi}_i^0(\overline{x}) = 0 \forall \overline{x} \in \overline{S}_i \setminus \{0\}$.

$$\overline{Q}_i = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ M \end{array} \begin{array}{c} \begin{matrix} States & 0 & 1 & 2 & \ldots & M \end{matrix} \\ \begin{pmatrix} -\lambda_i & \lambda_i & 0 & 0\ldots & 0 \\ \mu & -(\mu+\lambda_i) & \lambda_i & 0\ldots & 0 \\ \mu & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & -(\mu+\lambda_i) & \lambda_i \\ \mu & 0 & \ldots & 0 & -\mu \end{pmatrix} \end{array} \tag{8}$$

Using this generator matrix, we shall now solve for the steady state and transient probabilities for the lumped-CTMC and use these to calculate the expected steady state and transient aggregate net-rewards received by the students and the instructor (given by Theorem 6). Detailed proofs of the results presented in this section are provided in Appendix B.

6.3.1 Steady State Analysis

Steady state probability $\overline{\Pi}_i(\overline{x})$ is the probability that the CTMC $\overline{X}_i(t)$ (corresponding to student i) is in state $\overline{x}$, when $\overline{X}_i(t)$ has reached a steady state. Lemma 5 states that a steady state exists for this CTMC and so using the generator matrix $\overline{Q}_i$, we come up with the following equations for the steady state probability vector $\overline{\Pi}_i$ of $\overline{X}_i(t)$:

(Inflow in state $\overline{x}$ = Outflow from state $\overline{x}$) $\forall 0 \leq \overline{x} \leq M$

$$\sum_{\overline{y} \geq 1} \mu \overline{\Pi}_i(\overline{y}) = \lambda_i \overline{\Pi}_i(0), \text{ for } \overline{x} = 0 \tag{9}$$

$$\lambda_i \overline{\Pi}_i(\overline{x} - 1) = \lambda_i \overline{\Pi}_i(\overline{x}) + \mu \overline{\Pi}_i(\overline{x}), \forall \overline{x} : 0 < \overline{x} < M \tag{10}$$

$$\lambda_i \overline{\Pi}_i(M - 1) = \mu \overline{\Pi}_i(M), \text{ for } \overline{x} = M. \tag{11}$$

Lemma 6. *The solution to equations 9, 10, and 11 is:*

$$\overline{\Pi}_i(\overline{x}) = \frac{\rho_i^{\overline{x}}}{(1 + \rho_i)^{\overline{x}+1}}, \forall \overline{x} : 0 \leq \overline{x} < M, \text{ and}$$

$$\overline{\Pi}_i(M) = \frac{\rho_i^M}{(1 + \rho_i)^M}. \text{ where, } \rho_i = \frac{\lambda_i}{\mu}$$

Using these steady state probabilities, we calculate the expected steady state net-rewards of the students and the instructor (from student i's participation) in $\overline{X}_i(t)$.

Theorem 7. *The expected steady state net-rewards of the student i and the instructor in $\overline{X}_i(t)$ are given by:*

$$(a)\ \overline{R}_\Pi^{l,i} = \delta^{\log \mu} \rho_i \left(1 - \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l} \right) - \alpha_l \rho_i \left(1 - \left(\frac{\rho_i}{1 + \rho_i} \right)^M \right)$$

$$(b)\ \overline{R}_\Pi^{I,i} = \delta^{\log \mu} \rho_i \left(1 - \left(\frac{\rho_i}{1 + \rho_i} \right)^M \right) - \beta$$

6.3.2 Transient Analysis

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 $\overline{\pi}_i^t$'s of the lumped-CTMC $\overline{X}_i(t)$:

$$\frac{d\overline{\pi}_i^t(0)}{dt} = -\lambda_i \overline{\pi}_i^t(0) + \sum_{\overline{y}=1}^{M} \mu \overline{\pi}_i^t(\overline{y}) \tag{12}$$

$$\frac{d\overline{\pi}_i^t(\overline{x})}{dt} = -(\lambda_i + \mu)\overline{\pi}_i^t(\overline{x}) + \lambda_i \overline{\pi}_i^t(\overline{x} - 1), \forall \overline{x} : 0 < \overline{x} < M \tag{13}$$

$$\frac{d\overline{\pi}_i^t(M)}{dt} = -\mu \overline{\pi}_i^t(M) + \lambda_i \overline{\pi}_i^t(M - 1). \tag{14}$$

We now present some important results regarding transient probabilities. Detailed proofs of the results presented in this section are provided in Appendix C.

Lemma 7. *Given the initial distribution $\overline{\pi}_i^0$ for $\overline{X}_i(t)$ ($\overline{\pi}_i^0(0) = 1$ and $\overline{\pi}_i^0(\overline{x}) = 0\ \forall \overline{x} \in \{1, 2, \ldots, M\}$). The solution to the equations 12, 13, and 14 is given by:*

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

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

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

We now calculate the expected transient aggregate net-rewards over time T of a student i and the instructor (from student i's participation) using the lumped-CTMC $\overline{X}_i(t)$.

Theorem 8. *The expected transient aggregate net-rewards of the student i and the instructor in $\overline{X}_i(t)$ are given by:*

$$(a)\ \overline{R}_T^{l,i} = \delta^{\log \mu} \sum_{\overline{x}=0}^{m_l} (\overline{x} - m_l) \int_0^T \overline{\pi}_i^t(\overline{x}) dt + m_l \delta^{\log \mu} T - \alpha_l \sum_{\overline{x}=0}^{M} \overline{x} \int_0^T \overline{\pi}_i^t(\overline{x}) dt$$

$$(b)\ \overline{R}_T^{I,i} = \delta^{\log \mu} \sum_{\overline{x}=0}^{M} \overline{x} \int_0^T \overline{\pi}_i^t(\overline{x}) dt - \beta T. \tag{15}$$

By Theorem 6, we can calculate the expected net-rewards received by the students in an OEF from the net-reward obtained by the student i in the lumped-CTMC $\overline{X}_i(t)$. Thus we shall use the expected steady state reward $\overline{R}_\Pi^{l,i}$ and the expected transient aggregate reward $\overline{R}_T^{l,i}$ for student i of the lumped-CTMC $\overline{X}_i(t)$ to derive the expected steady-state $R_\Pi^{l,i}$ and transient aggregate net-rewards $R_T^{l,i}$ respectively received by the students in the OEF.

By Theorem 6, we can calculate the net-reward received by the instructor in an OEF by taking a weighted sum over the expected net-rewards to the instructor received through the lumped-CTMCs $\overline{X}_i(t)$s ($1 \le i \le n$). Thus we shall now compute the instructor's expected steady state net-reward R_Π^I and the expected transient aggregate net-reward R_T^I in the OEF by respectively taking weighted sums over the expected steady state net-rewards $\sum_{1 \le i \le n} c_i \overline{R}_\Pi^{I,i}$ and over the expected transient aggregate net-rewards $\sum_{1 \le i \le n} c_i \overline{R}_T^{I,i}$ the instructor receives in the lumped CTMCs $\overline{X}_i(t)$s.

7   OEF as a Stackelberg Game (Strategic Setting)

So far, we have assumed that the arrival rates of the instructor and the enrolled students are known to the model but this is unrealistic in the real world as the students and the instructor are strategic agents who will have a range of arrival rates to choose from and wish to maximize their utilities by choosing an optimal strategy (rate of arrival). We summarize important notations of this section in Table 3.

Symbol	Meaning
n	Number of students in the course.
L	Number of types of students.
T	Course Duration.
G^{ins}	Finite strategy set of the instructor (of size v).
G^{st}	Finite strategy set of the students (of size w).
Δ_a	A pure strategy strategy of the instructor.
Λ_b	A pure strategy strategy of a student .
ϕ_a	Proportion of times that pure strategy Δ_a is used.
ψ^i	Pure/mixed strategy over G^{st} of the student i.
$D_{a,b}^{\Pi,l,i}$	Expected steady state net-reward received by the student i of $Type_l$ if she chooses pure strategy Λ_b and instructor chooses the pure strategy Δ_a.
$D_{a,b}^{T,l,i}$	Expected transient aggregate net-reward (over time T) received by student i of $Type_l$ when she chooses pure strategy Λ_b and instructor chooses the pure strategy Δ_a.
$U_\Pi^{l,i}$	Expected steady state utility to a student i of $Type_l$.
U_T^I	Expected aggregate transient utility (over time T) to the instructor.
U_Π^I	Expected steady state utility to the instructor.
$B_{a,b}^{\Pi,I,i}$	Expected steady state utility received by the instructor if the student i chooses pure strategy Λ_b and the instructor chooses the pure strategy Δ_a.
$B_{a,b}^{T,I,i}$	Expected transient utility received by the instructor if the student i chooses pure strategy Λ_b and the instructor chooses the pure strategy Δ_a.

Table 3: Important notations in the Stackelberg game formulation

We assume that the instructor is a social-welfare maximizing agent whose goal is to maximize the students' understanding of the subject by enabling them to discuss frequently among themselves in the OEF. 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 (of visiting the forum and answering a question). We assume that the instructor's strategy will be announced first following which the students will decide their strategy.

We thus formulate an OEF as a Stackelberg game where the players are: an instructor offering the online course and each of the n students enrolled for the course. We make the following assumptions related to the formulation of the Stackelberg game.

- Instructor acts as a leader who decides her strategy (rate of arrival) first and the students act as the followers who after observing the instructor's strategy will finalize their own strategy (rate of arrival) in order to maximize their utility.
- 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) as derived in Theorems 7 and 8 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 putting these two together in a bi-level optimization problem which solves for the optimal instructor and students' strategies.

We establish some basic notations before we formulate the game. 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. These are the set of arrival rates from which the instructor and the students can choose their respective rates of arrival. The rates in the sets G^{ins} and G^{st} are bounded and are in the range $[0, \widehat{q}]$.

$$G^{ins} = \{\Delta_a | a \in \{1, \ldots, v\}, \Delta_a \in [0, \widehat{q}], \widehat{q} < \infty\}$$
$$G^{st} = \{\Lambda_b | b \in \{1, \ldots, w\}, \Lambda_b \in [0, \widehat{q}], \widehat{q} < \infty\}$$

The instructor can choose a pure/mixed strategy over G^{ins} and we denote the instructor's strategy by ϕ which is a point in probability simplex, $\Omega(G^{ins})$, where

$$\Omega(G^{ins}) = \{\phi = (\phi_1, \cdots, \phi_v) | 0 \leq \phi_a \leq 1, \sum_{a=1}^{v} \phi_a = 1, 1 \leq a \leq v\}$$

The value ϕ_a is the proportion of times that pure strategy Δ_a is used. Similarly, ψ^i denotes the pure/mixed strategy over G^{st} of the student i which is a point in probability simplex $\Omega(G^{st})$, where

$$\Omega(G^{st}) = \{\psi^i = (\psi_1^i, \cdots, \psi_w^i) | 0 \leq \psi_b^i \leq 1, \sum_{b=1}^{w} \psi_b^i = 1, 1 \leq b \leq w\}$$

The value ψ_b^i basically represents the proportion of times in which pure strategy Λ_b is used.

7.1 Student Optimization Problem

We define expected net-reward matrices $D^{\Pi,l,i}$ and $D^{T,l,i}$ corresponding to each student i of type l where each $D_{a,b}^{\Pi,l,i}$ entry denotes the expected steady state net-reward received by the student i if she chooses pure strategy Λ_b and instructor chooses the pure strategy Δ_a. Similarly $D_{a,b}^{T,l,i}$ denotes the expected transient aggregate net-reward (over time T) received by student i of $Type_l$ when she chooses pure strategy Λ_b and instructor chooses the pure strategy Δ_a.

Now, suppose the instructor has fixed a (pure/mixed) strategy ϕ and the student i has fixed a (pure/mixed) strategy ψ^i. Using the above definitions, the expected steady state utility $U_{\Pi}^{l,i}$ and the expected transient aggregate utility $U_T^{l,i}$ (over the course duration T) to a student i of $Type_l$ on the forum is defined by:

$$U_{\Pi}^{l,i} = \sum_{a=1}^{v} \sum_{b=1}^{w} D_{a,b}^{\Pi,l,i} \phi_a \psi_b^i; \quad U_T^{l,i} = \sum_{a=1}^{v} \sum_{b=1}^{w} D_{a,b}^{T,l,i} \phi_a \psi_b^i \qquad (16)$$

where, $D_{a,b}^{\Pi,l,i} = \overline{R}_{\overline{\Pi}}^{l,i}$ over the CTMC $\overline{X}_i(t)$(See Theorem 7 (a)),

$$D_{a,b}^{T,l,i} = \overline{R}_T^{l,i} \text{ over the CTMC } \overline{X}_i(t)\text{(See Theorem 8 (a)),}$$

and $\overline{X}_i(t)$ is defined by $\mu = \Delta_a$ and $\lambda_i = \Lambda_b$(See Equation 8).

20

Proposition 1. *Students i, j belonging to the same type $Type_l$ receive equal utility (steady state and transient aggregate) if they choose the same policy i.e. if $\psi^i = \psi^j (= \psi^l)$ for students $i, j \in Type_l$ then,*

$$U_\Pi^{l,i} = U_\Pi^{l,j}, \quad U_T^{l,i} = U_T^{l,j}.$$

Proof. Please check the Appendix D section of the paper for the detailed proof of this result. □

Using the above definitions, we now formulate the (steady state) optimization problem for student i of $Type_l$ when the instructor has fixed a strategy ϕ.

$$\psi^{i*} = arg\max_{\psi^i} \sum_{a=1}^{v} \sum_{b=1}^{w} D_{a,b}^{\Pi,l,i} \phi_a \psi_b^i, \quad \text{s.t.} \sum_{b=1}^{w} \psi_b^i = 1, \ \psi_b^i \geq 0. \tag{17}$$

Note that ψ^{i*} is the optimal rate of arrival of student i of $Type_l$ in response to instructor strategy ϕ.

Note that students belonging to the same type can have different optimal strategies in response to the same instructor strategy as the optimization problem given in Equation 17 can have multiple solutions. Due to symmetry among students belonging to the same type, we assume that students belonging to the same type will choose the same optimal strategy i.e. $\psi^{i*} = \psi^{j*} = \psi^{l*} \ \forall i, j \in Type_l$.

Hence, we can infer that, instead of solving the student (steady state) optimization problem (Equation 17) for each student i, j belonging to a particular type $Type_l$, we can just solve the student optimization problem for a single representative $Type_l$ student (now represented as l) and her optimal policy ψ^{l*} would be followed by each student i of $Type_l$. Thus the (steady state) optimization problem to be solved by each representative student l ($1 \leq l \leq L$) will be :

$$\psi^{l*} = arg\max_{\psi^l} \sum_{a=1}^{v} \sum_{b=1}^{w} D_{a,b}^{\Pi,l} \phi_a \psi_b^l, \quad \text{s.t.} \sum_{b=1}^{w} \psi_b^l = 1, \ \psi_b^l \geq 0 \tag{18}$$

where, $D_{a,b}^{\Pi,l} = D_{a,b}^{\Pi,l,i}$ for any $i \in Type_l$.

The optimization problem for the expected transient aggregate utilities will be similar to the steady-state counterpart discussed above.

7.2 Instructor Optimization Problem

We now define similarly the steady state utility U_Π^I and the aggregate transient utility U_T^I (over time T) to the instructor when she has fixed her strategy as ϕ and the n students have fixed their policies as ψ^i ($1 \leq i \leq n$). We define net-reward matrices $B^{\Pi,I,i}$ and $B^{T,I,i}$ for the instructor corresponding to each student i where each $B_{a,b}^{\Pi,I,i}$ entry denotes the steady state utility received by the instructor and each $B_{a,b}^{T,I,i}$ entry denotes the aggregate utility (over time T) received by the instructor w.r.t. student i's arrivals, if the student i chooses pure strategy Λ_b and the instructor chooses the pure strategy Δ_a.

$$U_\Pi^I = \sum_{a=1}^{v} \sum_{b=1}^{w} \sum_{i=1}^{n} B_{a,b}^{\Pi,I,i} \phi_a \psi_b^i; \quad U_T^I = \sum_{a=1}^{v} \sum_{b=1}^{w} \sum_{i=1}^{n} B_{a,b}^{T,I,i} \phi_a \psi_b^i \tag{19}$$

where, $B_{a,b}^{\Pi,I,i} = c_i \overline{R}_\Pi^{I,i}$ over the CTMC $\overline{X}_i(t)$ (See Theorem 7 (b)),

$$B_{a,b}^{T,I,i} = c_i \overline{R}_T^{I,i} \text{ over the CTMC } \overline{X}_i(t) \text{ (See Theorem 8 (b))},$$

and $\overline{X}_i(t)$ is defined by $\mu = \Delta_a$ and $\lambda_i = \Lambda_b$. (See Equation 8).

Proposition 2. *The net-reward matrices $B^{\Pi,I,i}, B^{\Pi,I,j}$ (steady state) and $B^{T,I,i}, B^{T,I,j}$ (transient aggregate) for the instructor w.r.t. students i, j belonging to $Type_l$ have the following property:*

$$B^{\Pi,I,i} = B^{\Pi,I,j} (= B^{\Pi,I,l}), \quad B^{T,I,i} = B^{T,I,j} (= B^{T,I,l}).$$

Proof. Please check the Appendix D section of the paper for the detailed proof of this result. □

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 such that she can maximize her utility. The instructor will maximize her utility by choosing a solution ϕ^* to the following (steady state) optimization problem:

$$\phi^* = \arg\max_{\phi} \sum_{a=1}^{v} \sum_{b=1}^{w} \sum_{i=1}^{n} B_{a,b}^{\Pi,I,i} \phi_a [\psi^{i*}(\phi)]_b \tag{20}$$

$$s.t. \sum_{a=1}^{v} \phi_a = 1, \quad \phi_a \in [0 \dots 1]$$

$\psi^{i*}(\phi)$ is the optimal strategy (See Equation 17) of student i when ϕ is the instructor policy. We know that there are n_l students belonging to $Type_l$. As given in previous section, we know that $\psi^{i*}(\phi) = \psi^{j*}(\phi) = \psi^{l*}(\phi) \ \forall i, j \in Type_l$. Using, Proposition 2, we get:

$$\phi^* = \arg\max_{\phi} \sum_{a=1}^{v} \sum_{b=1}^{w} \sum_{l=1}^{L} n_l B_{a,b}^{\Pi,I,l} \phi_a [\psi^{l*}(\phi)]_b$$

$$s.t. \sum_{a=1}^{v} \phi_a = 1, \quad \phi_a \in [0 \dots 1]$$

Dividing the objective function of this (steady state) optimization problem by n, we get

$$\phi^* = \arg\max_{\phi} \sum_{a=1}^{v} \sum_{b=1}^{w} \sum_{l=1}^{L} p^l B_{a,b}^{\Pi,I,l} \phi_a [\psi^{l*}(\phi)]_b$$

$$s.t. \sum_{a=1}^{v} \phi_a = 1, \quad \phi_a \in [0 \dots 1]$$

Here, $p^l = n_l/n$ i.e. the proportion of the class that belongs to $Type_l$. Using the student optimization problem in Equation 18, we can rewrite the above problem as:

$$\phi^* = \arg\max_{\phi} \sum_{a=1}^{v} \sum_{b=1}^{w} \sum_{l=1}^{L} p^l B_{a,b}^{\Pi,I,l} \phi_a \psi_b^{l*} \tag{21}$$

$$s.t., \psi^{l*} = \arg\max_{\psi^l} \sum_{a=1}^{v} \sum_{b=1}^{w} D_{a,b}^{\Pi,l} \phi_a \psi_b^l,$$

$$\sum_{a=1}^{v} \phi_a = 1, \quad \phi_a \in [0 \dots 1], \quad \sum_{b=1}^{w} \psi_b^l = 1, \quad \psi_b^l \geq 0.$$

Note that the size of the optimization problem reduces considerably with the introduction of the similarity among students belonging to the same type. Now we have to solve the optimization problem w.r.t to only $L + 1$ agents instead of $n + 1$ agents and in general $L << n$. The optimization problem using the expected transient aggregate utilities for the instructor will be similar to the steady-state counterpart discussed above.

7.3 MILP Formulation

The (steady state) optimization problem given in Equation 21 is essentially a mixed integer quadratic program (MIQP). We convert this MIQP to a mixed integer linear program (MILP) by following the approach in [Paruchuri et al., 2008] where they solve a Bayesian Stackelberg game by converting a similar bi-level optimization problem into an MILP (See **Proposition 2** in [Paruchuri et al., 2008]). Thus, we get the following:

$$\max_{\psi,\xi,\eta} \sum_{l=1}^{L} \sum_{a=1}^{v} \sum_{b=1}^{w} p^l B_{a,b}^{\Pi,I,l} \xi_{ab}^l \tag{22}$$

$$s.t. \sum_{a=1}^{v} \sum_{b=1}^{w} \xi_{ab}^l = 1, \forall l \; ; \; \sum_{b=1}^{w} \xi_{ab}^l \leq 1, \forall l$$

$$\psi_b^l \leq \sum_{a=1}^{v} \xi_{ab}^l \leq 1, \forall l, \forall b \; ; \; \sum_{b=1}^{w} \psi_b^l = 1, \forall l$$

$$0 \leq (\eta^l - \sum_{a=1}^{v} D_{ab}^{\Pi,l}(\sum_{h=1}^{w} \xi_{ah}^l)) \leq (1 - \psi_b^l)N, \forall l, \forall b \; ; \; \eta \in \mathbb{R}$$

$$\sum_{b=1}^{w} \xi_{ab}^l = \sum_{b=1}^{w} \xi_{ab}^1, \forall l, \forall a \; ; \; \psi_b^l \in \{0,1\}\forall l \; ; \; \xi_{ab}^l \in [0\ldots1]\forall l$$

Here, $\xi_{ab}^l = \phi_a \psi_b^l$, η is derived through duality, and N is a large constant (The details can be seen in Parachuri et. al [Paruchuri et al., 2008]).

Analogously, we can come up with MILP formulation considering the expected transient aggregate utilities and they are given below.

$$\max_{\psi,\xi,\eta} \sum_{l=1}^{L} \sum_{a=1}^{v} \sum_{b=1}^{w} p^l B_{a,b}^{T,I,l} \xi_{ab}^l \tag{23}$$

$$s.t. \sum_{a=1}^{v} \sum_{b=1}^{w} \xi_{ab}^l = 1, \forall l \; ; \; \sum_{b=1}^{w} \xi_{ab}^l \leq 1, \forall l$$

$$\psi_b^l \leq \sum_{a=1}^{v} \xi_{ab}^l \leq 1, \forall l, \forall b \; ; \; \sum_{b=1}^{w} \psi_b^l = 1, \forall l$$

$$0 \leq (\eta^l - \sum_{a=1}^{v} D_{ab}^{T,l}(\sum_{h=1}^{w} \xi_{ah}^l)) \leq (1 - \psi_b^l)N, \forall l, \forall b \; ; \; \eta \in \mathbb{R}$$

$$\sum_{b=1}^{w} \xi_{ab}^l = \sum_{b=1}^{w} \xi_{ab}^1, \forall l, \forall a \; ; \; \psi_b^l \in \{0,1\}\forall l \; ; \; \xi_{ab}^l \in [0\ldots1]\forall l$$

8 Simulations

The solution to the optimization problem (22) (based on steady state utilities) and optimization problem (23) (based on transient utilities) provide us with the optimal strategies ϕ^* and ψ^{i*}'s for the instructor and the students respectively. 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 the different parameters of the model. Table 4 lists the values and ranges for various parameters used in the simulations.

8.1 OEF-CTMC: Transient Vs Steady State

We substantiate the importance of transient analysis [Narahari and Viswanadham, 1994] of the proposed CTMC for online classroom setting by studying the variation of the normalized error between the student utility values obtained through the steady and transient analyses. We define

$$\text{Normalized Error} = \left| \frac{U_{\Pi}^{l,i} - \frac{U_T^{l,i}}{T}}{U_{\Pi}^{l,i}} \right|, \text{ where } U_{\Pi}^{l,i} \text{ and } U_T^{l,i} \text{ are respectively}$$

the steady state and transient utilities of a student as defined in Equation 16.

For illustration purposes, we assume that only $Type_1$ students are present in the classroom and fix the max-reward and cost per arrival for each of these students as: $m = 3$, $\alpha = 0.3$. We observe the variation in the normalized error with time (Figure 8).

Parameter	Definition	Range/Value (For Model)	Range/Value (Fixed for Simulations)
n	Total no. of students in the online classroom	Any finite value	1000
M	Maximum number of answers from a student (Last state of the lumped CTMC)	Any finite value	100
L	Number of types of students being considered	Any finite value ($\leq n$)	2
$Type_l$	Set containing the students of Type l	Finite set of students	$|Type_1| = |Type_2| = n/2$
T	Duration of the online course (Units: Weeks)	Any finite value	11
λ	Rate of arrival of **one** student (per Week)	Any finite set of values	$\{1, 4, 5, 6, 7, 8, 10\}$
μ	Rate of arrival of the instructor (per Week)	Any finite set of values	$\{1, 2, \ldots, 10\}$
α_l ($1 \leq l \leq L$)	Cost per arrival to a $Type_l$ student	$(0, 1)$	$\alpha_1 \in \{0.1, 0.2, 0.3\}$ ($Type_1$ - High Caliber/Low Cost) $\alpha_2 \in \{0.6, 0.8, 0.9\}$ ($Type_2$ - Low Caliber/High Cost)
β	Cost per arrival to the instructor	$(0, 1)$	0.5
δ	Willingness of the instructor to reward	$(0, 1)$	0.9
m_l	Maximum number of rewards per question for each student $i \in Type_l$	Any finite value	$\{2, 6, 10\}$
c_i	Bias of the instructor towards a student i	$c_i \in (0, 1), \sum_{i=1}^{n} c_i = 1$ $c_i = c_j \; \forall i, j \in Type_l$	Extreme Bias towards $Type_1$ (or, $Type_2$) students: $c_i = \frac{0.99}{n_1} \left(\text{or, } \frac{0.01}{n_1}\right) \forall i \in Type_1,$

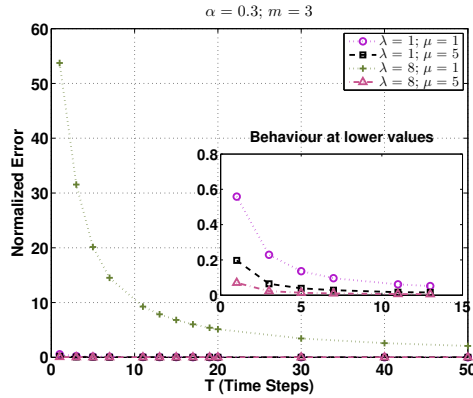Table 4: Parameters of the Model

Figure 8:   Study normalized error in student utility using the steady state and transient analyses as time evolves.

We observe from Figure 8 that the normalized error tends to zero as time progresses. But this error becomes negligible only after 10-15 weeks and is quite significant before that period. The typical duration of online classrooms is around 4-10 weeks [Haggard, 2013] and this necessitates the use of transient analysis over steady state analysis in such a setting for accurate results. Note that the normalized error can be any value greater than or equal to zero as the steady state and avg. aggregate utilities can have the same or opposite signs at a time t. Also, note that we observe a similar behavior for other parameter configurations as well and hence we do not discuss them here. Also, due to the significance of the transient analysis, we shall provide the simulation results from only the transient analysis of our model. The results based on steady state analysis follow a similar trend.

8.2   Heterogeneous Student Participation: Effect of Budget and Instructor Rate

We study the variation in the optimal student participation rate (λ^*) of the heterogeneous student population with μ and m for different student types in Figure 9. We observe that, if we keep μ constant and increase the reward m, then λ^* either increases or remains the same. Hence, giving higher rewards can increase student participation but only up to a certain extent.

If we fix m and study the variation of λ^* with changing μ in Figure 9(a) to Figure 9(d), we observe that as the cost per arrival (α) of a student increases, λ^* becomes sensitive to μ. This means that the students with lower α are self motivated and have only a limited impact of μ (See Figure 9(a)). They act as super-posters and aggressively post on the OEF being only slightly affected by the instructor participation rates. This phenomenon is also corroborated in our empirical study as shown in Figure 3(a).
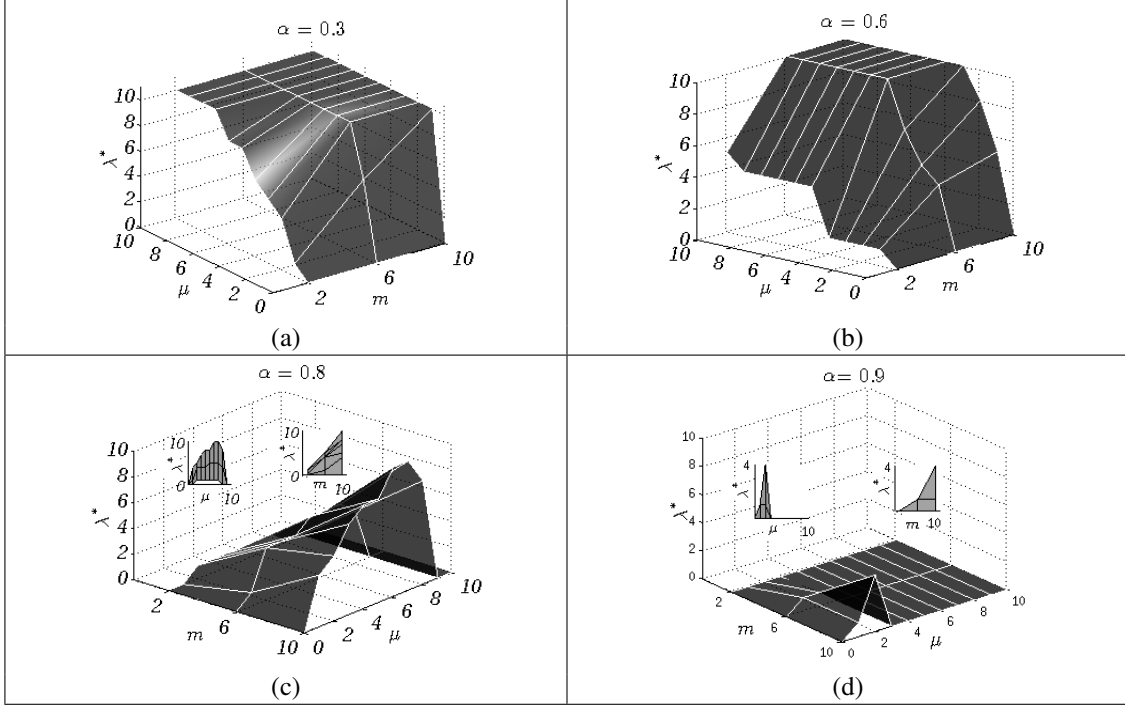

Figure 9: Variation of the optimal participation rates (λ^*) of each student belonging to the different types (characterized by four different α's) with change in the instructor's arrival rate (μ) and the maximum number of rewards (m) given to that type. The $X - Z$ and $Y - Z$ projections of Figures 9c and 9d are given as insets to aid the understanding of the graphs.

However, we observe non-monotonic participation patterns for students with high cost per arrival ($\alpha = 0.8, 0.9$) in Figures 9(c) and 9(d) as λ^* initially increases with increasing μ and then, with any further increase in μ, λ^* starts falling. This trend has been noted in literature (for example: [Ghosh and Kleinberg, 2013]). Our model is thus able to corroborate real-world behavior of heterogeneous students in an OEF varying from the disinterested students to those who are highly active. We observe that the active students are self-motivated and will visit the forum even for low rewards and their level of participation generally increases as the instructor arrives more frequently on the OEF. However, the not-so-active students want high rewards and lower instructor participation i.e., they want lesser questions (and more time) with higher incentives.

8.3 Instructor-Student Arrival Rates: Effect of Instructor Bias

It is generally observed that instructors have varying notions of bias towards different types of students. Some instructors may value the participation from the weaker students higher than good students while others may focus on the brighter students. In our model, the parameter c_i (See Table 4) captures the inherent instructor bias towards a student i. This parameter is an input parameter to the model and is set based on how the instructor values an answer from the different types of students.

We consider a class with 1000 students categorized into two types (Table 4). For illustration purposes, we consider 2 students from this class: Student 1 (belonging to $Type_1$ with low α_1) who is very active on the forum and Student 2 (belonging to $Type_2$ with high α_2) who does not like to post frequently on the OEF. We fix the following 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 Student 1 and as a high value ($c_2 = (0.99/n_2)$) for Student 2 where n_1 and n_2 are number of students of $Type_1$ and $Type_2$ respectively.
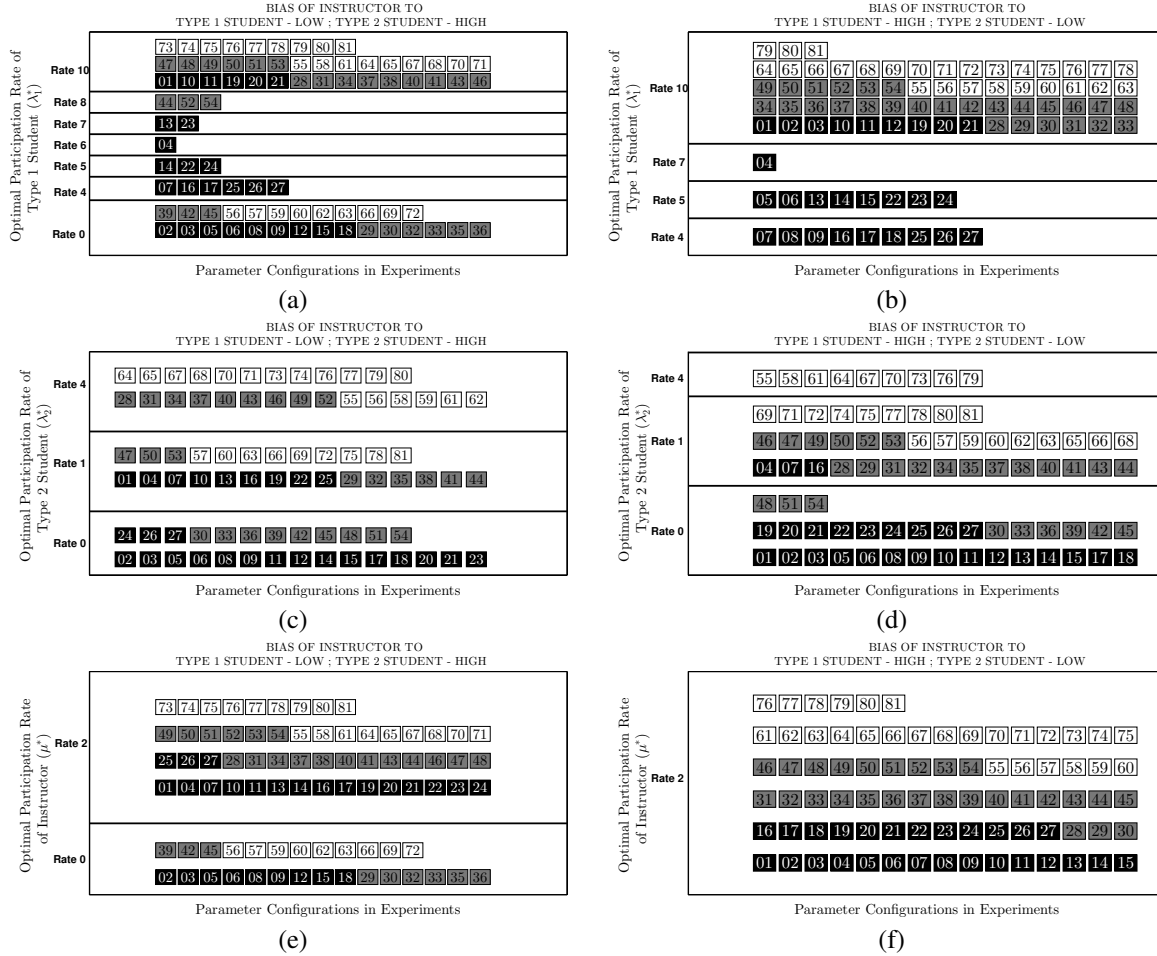
Figure 10: Effect of Instructor bias on the optimal participation rates of a Type 1 student. Each square represents the respective optimal participation rate of a Type 1 student depending on parameter configuration in experiment. Black, Grey and White squares indicate configurations where $m_1 = 2$(low rewards), $m_1 = 6$ (medium rewards) and $m_1 = 10$ (high rewards) respectively.

We examine the optimal rates of participation for Student 1, Student 2, and instructor for each of the possible 81 configurations in the presence of instructor bias towards certain type of students. Note that a configuration basically corresponds to a fixed value for the parameters: (α_1, α_2, m_1, m_2).

We will initially consider an instructor with low bias towards Student 1 and very high bias to Student 2 i.e., we set $c_1 = (0.01/n_1)$ and $c_2 = (0.99/n_2)$ where n_1 and n_2 are the number of students of $Type_1$ and $Type_2$ respectively and c_i is the bias of instructor towards Student i. Note that these parameters affect the utility values of the instructor as described in Equation 19. Under this setting, we generate scatter plots depicting optimal participation rates where each point in the scatter plot corresponds to one of the 81 parameter configurations. These 81 parameter configurations are numbered uniquely so that we can compare the results when the instructor reverses his bias values. The scatter plots are given in Figure 10(a), Figure 10(c), and Figure 10(e) and these correspond to the optimal rates of participation for Student 1, Student 2, and instructor respectively. Note that these plots correspond to the scenario when the instructor has low bias to Student 1 and high bias to Student 2. We change the instructor behaviour to have high bias towards Student 1 and very low bias towards Student 2 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 participation rates of Student 1, Student 2, and instructor are given in Figure 10(b), Figure 10(d), and Figure 10(f) 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 10(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 10(a) have optimal rate 10 while config 09 improved from rate 0 in Figure 10(a) to rate 4 in Figure 10(b)). We also see that there is a dip in participation of Student 2 for many configurations when the instructor decreases his bias to Student 2 (See Figure 10(c) and Figure 10(d)). For example, see configs 01, 28 in

Figure 10(c) and Figure 10(d). In Figure 10(c) for Config 01, Student 2 participates with Rate 1 whereas when the instructor decreases bias to Student 2, for this configuration, Student 2 optimal rate decreases to Rate 0 as shown in Figure 10(d). Similarly Student 2 decreases his optimal rate from Rate 4 to Rate 2 when the instructor decreases the bias for Student 2 for the configuration 28 as shown in Figure 10(c) and Figure 10(d).

We also observe that the instructor needs to increase his/her participation when there is a higher bias to Student 1 than the scenario when there is a higher bias to Student 2. This is beacuse Student 1 is an enthusiastic student compared to Student 2 and hence, to keep the attention of Student 1 , the instructor needs to post more questions on the forum. In the scenario where the instructor has high bias to Student 2, the optimal rate of the instructor can be low as Student 2 dislikes answering too many questions on the forum as the cost of answering the questions overshadows the rewards obtained by Student 2.

To summarize, these experiments shed interesting insights into the role of instructor bias parameter of our proposed OEF model and thus, makes our proposed OEF model more realistic. It has been observed that participation of students is indeed affected by instructor biases towards certain students and the instructor bias parameter enables us to capture this effect in our proposed model for OEFs.

9   Conclusions

In this work, we performed empirical analysis of online education forums (OEFs) of the Game Theory course offered in the Indian Institute of Science in 2012 and 2014. We identified several key parameters which dictate the activities of OEFs like student heterogeneity, effective incentive design, super-posters phenomenon, etc. Motivated by empirical observations, we proposed a continuous time Markov chain (CTMC) model to capture instructor -student interactions in an OEF. Using concepts from lumpability of CTMCs, we performed steady state and transient analysis evaluate expected steady state and transient aggregate net-rewards for the instructor and the students. We then formulated a mixed-integer linear program which views the OEFs strategically as a single-leader-multiple-follower Stackelberg game. We undertook detailed simulations and develop new insights into the activities of OEF by studying the effects of parameters like participation costs, budget and instructor bias on the student/instructor participation rates. Our model corroborated with some key empirical observations and could recommend an optimal plan to the instructor for maximizing student participation in OEFs.

10   Future Work

The analysis used in the paper assumed certain parameters (for eg: m_l parameter in Table 1) to be fixed. An additional preprocessing stage could be introduced to optimally solve for such parameters which could improve the results of the model. Further, students could collude on their strategies to improve utilities while reducing their participation levels. Devising instructor-driven collusion-resistant strategies could be an interesting direction. Our work does not consider effort modeling of the students. Information about the effort put by the students could guide the incentive design and may result in better quality of participation among the students in the OEFs.

References

Coursera. Coursera mooc platform, 2017. URL https://www.coursera.org/about/community.

Alexandru Butean, Popa Emil, and Mihai Stancu. An approach to improve elearning platforms accessibility. In *Conference proceedings of eLearning and Software for Education (eLSE)*, pages 278–284. Universitatea Nationala de Aparare Carol I, 2015.

Barbara Means, Yukie Toyama, Robert Murphy, Marianne Bakia, and Karla Jones. Evaluation of evidence-based practices in online learning. a meta-analysis and review of online learning studies. Technical report, U.S. Department of Education,, 2009.

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

Louise F Pendry and Jessica Salvatore. Individual and social benefits of online discussion forums. *Computers in Human Behavior*, 50:211–220, 2015.

Margie Martyn. The hybrid online model: Good practice. *Educause Quarterly*, 26(1):18–23, 2003.

R.M. Palloff and K. Pratt. *Building Online Learning Communities: Effective Strategies for the Virtual Classroom*. The Jossey-Bass Higher and Adult Education Series. Wiley, 2007. ISBN 9780787988258. URL http://books.google.co.in/books?id=VSnL3mla_x8C.

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

L Hopperton. Computer conferencing and college education. *The College Quarterly*, 5(2), 1998.

Sherry L Markel. Technology and education online discussion forums: It's in the response. *Online journal of distance learning administration*, 4(2), 2001.

Arpita Ghosh and Preston McAfee. Crowdsourcing with endogenous entry. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pages 999–1008, 2012.

Arpita Ghosh and Patrick Hummel. Implementing optimal outcomes in social computing: A game-theoretic approach. *CoRR*, abs/1202.3480, 2012.

Yuko Sakurai, Tenda Okimoto, Masaaki Oka, Masato Shinoda, and Maokoto Yokoo. Quality-control mechanism utilizing worker's confidence for crowdsourced tasks. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, AAMAS '13, pages 1347–1348, 2013.

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

Lori Breslow, David E Pritchard, Jennifer DeBoer, Glenda S Stump, Andrew D Ho, and Daniel T Seaton. Studying learning in the worldwide classroom: Research into edx's first mooc. *Research and Practice in Assessment*, 8, 2013.

Joseph Jay Williams, René F Kizilcec, Daniel M Russell, and Scott R Klemmer. Learning innovation at scale. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 131–134. ACM, 2014.

Janice Ballou. *Open-Ended Question*, pages 548–550. SAGE Publications, Inc., 0 edition, 2008.

Darrell M Hull and Terrill F Saxon. Negotiation of meaning and co-construction of knowledge: An experimental analysis of asynchronous online instruction. *Computers & Education*, 52(3):624–639, 2009.

Jennifer DeBoer, Glenda S Stump, Daniel Seaton, and Lori Breslow. Diversity in mooc students' backgrounds and behaviors in relationship to performance in 6.002 x. In *Proceedings of the Sixth Learning International Networks Consortium Conference*, volume 4, 2013.

René F Kizilcec, Chris Piech, and Emily Schneider. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pages 170–179. ACM, 2013.

Julia Wilkowski, Amit Deutsch, and Daniel M. Russell. Student skill and goal achievement in the mapping with google mooc. In *ACM Learning at Scale*, 2014.

Geoffrey A. Fowler. An early report card on massive open online courses. *Wall Street Journal*, 1, Oct 8 2013.

Sherif Halawa, Daniel Greene, and John Mitchell. Dropout prediction in moocs using learner activity features. *eLearning Papers*, 37:1–10, 02/2014 2014.

Justin Cheng, Chinmay Kulkarni, and Scott Klemmer. Tools for predicting drop-off in large online classes. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*, pages 121–124. ACM, 2013.

Ronen I Brafman and Moshe Tennenholtz. On partially controlled multi-agent systems. *Journal of Artificial Intelligence Research*, 4:477–507, 1996.

Francesco Maria Delle Fave, Albert Xin Jiang, Zhengyu Yin, Chao Zhang, Milind Tambe, Sarit Kraus, and John P Sullivan. Game-theoretic patrolling with dynamic execution uncertainty and a case study on a real transit system. *Journal of Artificial Intelligence Research*, 50:321–367, 2014.

Alejandro Uriel Carbonara, Anupam Datta, Arunesh Sinha, and Yair Zick. Incentivizing peer grading in moocs: An audit game approach. In *IJCAI*, 2015.

Ganesh Ghalme, Sujit Gujar, Amleshwar Kumar, Shweta Jain, and Y. Narahari. Design of coalition resistant credit score functions for online discussion forums. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '18, page 95–103, Richland, SC, 2018. International Foundation for Autonomous Agents and Multiagent Systems.

Piazza. Piazza online gathering forum, 2017.

Praveen Paruchuri, Jonathan P Pearce, Janusz Marecki, Milind Tambe, Fernando Ordonez, and Sarit Kraus. 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*, pages 895–902, 2008.

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

John G Kemeny and James Laurie Snell. *Finite markov chains*, volume 356. van Nostrand Princeton, NJ, 1960.

Peter Buchholz. Exact and ordinary lumpability in finite Markov chains. *J. Appl. Prob.*, 31(1):59–75, 1994. ISSN 00219002.

Salem Derisavi, Holger Hermanns, and William H Sanders. Optimal state-space lumping in markov chains. *Information Processing Letters*, 87(6):309–315, 2003.

Jianjun Paul Tian and D. Kannan. Lumpability and commutativity of markov processes. *Stochastic Analysis and Applications*, 24(3):685–702, 2006.

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

Y Narahari and N Viswanadham. Transient analysis of manufacturing systems performance. *IEEE Transactions on Robotics & Automation*, 10(2):230–244, 1994.

Stephen Haggard. The maturing of the mooc: Literature review of massive open online courses and other forms of online distance learning. Technical report, Media Legislation Reports, September 2013.

T. Tao. *An Introduction to Measure Theory*. Graduate studies in mathematics. American Mathematical Soc., 2011.

N. Viswanadham and Y. Narahari. *Performance Modeling of Automated Manufacturing Systems*. Prentice Hall PTR, 1992.

Robert G Gallager. *Discrete stochastic processes*, volume 321. Springer Science & Business Media, 2012.

A   Proofs : Lumpability

Theorem 5. *(i)* $X(t) = (\mathcal{S}, Q)$ *is lumpable w.r.t. partition* $\overline{S}_i = \{\overline{S}_i^a | a \in \{0, 1, \dots, M\}\}$.

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

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

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

Proof. As given in Definition 1, $X(t)$ 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. Fig. 11 shows the two blocks $\overline{S}_i^a$ and $\overline{S}_i^b$.
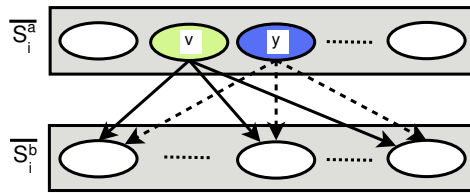
Figure 11: Two blocks S_i^a and S_i^b of the partition S_i are shown. The solid lines depict transition from state v in the S_i^a block to states in the S_i^b block. And the dotted lines depict transition from state y in the S_i^a block to states in the S_i^b block. We need to prove that the sum of rates of solid transition lines is equal to the sum of rates of dotted transition lines.

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 $z \in S_i^b$ 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 $v, y \in \overline{S}_i^a$,

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

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

We consider the following exhaustive cases.

1. $\underline{b \neq a}$

   - $\underline{b = a + 1}$

     LHS $= \displaystyle\sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, v_{-i}), (b, z_{-i})) = \sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, v_{-i}), (a + 1, z_{-i}))$

     $= Q((a, v_{-i}), (a + 1, v_{-i})) + \displaystyle\sum_{z_{-i} \in \mathcal{S}_{-i} \backslash \{v_{-i}\}} Q((a, v_{-i}), (a + 1, z_{-i})) = \lambda_i + 0 = \lambda_i$

     (As the first Q expression satisfies Case (i) and all the other Q expressions in the summation satisfy Case (iv) of Equation 1.)

     RHS $= \displaystyle\sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, y_{-i}), (b, z_{-i})) = \sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, y_{-i}), (a + 1, z_{-i}))$

     $= Q((a, y_{-i}), (a + 1, y_{-i})) + \displaystyle\sum_{z_{-i} \in \mathcal{S}_{-i} \backslash \{y_{-i}\}} Q((a, y_{-i}), (a + 1, z_{-i})) = \lambda_i + 0 = \lambda_i$

     (As the first Q expression satisfies Case (i) and all the other Q expressions in the summation satisfy Case (iv) of Equation 1.)

     $\therefore LHS = RHS$ in this case.
   - $\underline{b > a + 1}$
     Similar analysis as above yields $LHS = RHS = 0$ as Case (iv) of Equation 1 is satisfied for the whole summation.
   - $\underline{b < a, b \neq 0}$
     Similar analysis as above yields $LHS = RHS = 0$ as Case (iv) of Equation 1 is satisfied for the whole summation.
   - $\underline{b = 0, b \neq a}$

     LHS $= \displaystyle\sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, v_{-i}), (b, z_{-i})) = \sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, v_{-i}), (0, z_{-i}))$

     $= Q((a, v_{-i}), (0, 0, \ldots, 0)) + \displaystyle\sum_{z_{-i} \in \mathcal{S}_{-i} \backslash \{(0, \ldots, 0)\}} Q((a, v_{-i}), (0, z_{-i})) = \mu + 0 = \mu$

     (As the first Q expression satisfies Case (ii) of Equation 1 and all the other Q expressions in the summation satisfy Case (iv) of Equation 1.)

     RHS $= \displaystyle\sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, y_{-i}), (b, z_{-i})) = \sum_{z_{-i} \in \mathcal{S}_{-i}} Q((a, y_{-i}), (0, z_{-i}))$

     $= Q((a, y_{-i}), (0, 0, \ldots, 0)) + \displaystyle\sum_{z_{-i} \in \mathcal{S}_{-i} \backslash \{(0, \ldots, 0)\}} Q((a, y_{-i}), (0, z_{-i})) = \mu + 0 = \mu$

     (As the first Q expression satisfies Case (ii) of Equation 1 and all the other Q expressions in the summation satisfy Case (iv) of Equation 1.)

     $\therefore LHS = RHS$ in this case.

2. $\underline{b = a}$

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

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

   (Last Q expression satisfy Case (iii) of Equation 1.)

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

   (Expanding the second summand)

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

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

   Similarly on solving for RHS we get,

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

   Now, from **Case** 1 we see that,

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

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

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

   $$\therefore LHS = RHS.$$

Thus we prove that $\forall \overline{S}_i^a, \overline{S}_i^b \in \overline{S}_i$ and for any $v, y \in \overline{S}_i^a$, $q(v, \overline{S}_i^b) = q(y, \overline{S}_i^b)$, thus the CTMC $X(t)$ is lumpable w.r.t partition $\overline{S}_i$. $\qquad \square$

Proof. (ii) Also, as stated in Definition 1(ii), the generator matrix for the quotient Markov chain $\overline{X}_i(t) = (\overline{S}_i, \overline{Q}_i)$ will be determined by $\overline{Q}_i(\overline{S}_i^a, \overline{S}_i^b) = q(x, \overline{S}_i^b)$ for any $\overline{S}_i^a, \overline{S}_i^b \in \overline{S}_i$, where x is any state in $\overline{S}_i^a$. And these quantities have already been computed in the Cases 1 and 2 above. Hence Proved. $\qquad \square$

Lemma 1. $\overline{\pi}_i^t(\overline{x}) = \sum_{x \in \overline{S}_i^{\overline{x}}} \pi^t(x)$

Proof. 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. Also because of the nature of the problem, the arrival rate λ_i for a student i ($i \in \{1, 2, \dots, n\}$) and the arrival of the instructor μ will be some finite values. Thus each value in the generator matrix Q of $X(t)$ defined in Equation 1 will be a finite value. Following a similar argument we can say that each value in the generator matrix $\overline{Q}_i$ of $\overline{X}_i(t)$ (for any $i \in \{1, 2, \dots, n\}$) defined in Equation 1 will be a finite value. Thus there exists a finite number $\widehat{q} < \infty$ which bounds the rate entries in the rate matrices Q and $\overline{Q}_i$ ($\forall i \in \{1, 2, \dots, n\}$).

Thus using Theorem 2 for CTMCs $X(t)$ and $\overline{X}_i(t)$ we can say that:

1. (a) there exists an underlying DTMC Y_k for the CTMC $X(t)$ with transition matrix $P = Q/\widehat{q} + I$.

(b) there exists an underlying DTMC $\overline{Y}_k^i$ for the CTMC $\overline{X}_i(t)$ with transition matrix $\overline{P}_i = \overline{Q}_i/\widehat{q} + I$.

2. (a) the distribution vector π^t of the CTMC $X(t)$ at time t starting with π^0 at time 0 is given by $\pi^t = \sum_{k=0}^{\infty} \exp(-\widehat{q}t) \left[(\widehat{q}t)^k/k!\right] \pi^k$, where π^k is the distribution vector of Y_k after k jumps.

   (b) the distribution vector $\overline{\pi}_i^t$ of the CTMC $\overline{X}_i(t)$ at time t starting with $\overline{\pi}_i^0$ at time 0 is given by $\overline{\pi}_i^t = \sum_{k=0}^{\infty} \exp(-\widehat{q}t) \left[(\widehat{q}t)^k/k!\right] \overline{\pi}_i^k$, where $\overline{\pi}_i^k$ is the distribution vector of $\overline{Y}_k^i$ after k jumps.

3. $X(t)$ is lumpable on Q w.r.t. the partition $\overline{S}_i$ (Theorem 5) thus Y_k is also lumpable on P w.r.t the partition $\overline{S}_i$.

Also, 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 $\pi^0(0,0,\ldots,0) = 1$, $\pi^0(y) = 0 \forall y \in \mathcal{S} \setminus \{(0,0,\ldots,0)\}$, $\overline{\pi}_i^0(0) = 1$, and $\overline{\pi}_i^0(\overline{x}) = 0 \forall \overline{x} \in \overline{S}_i \setminus \{0\}$. Note that $\overline{\pi}_i^0(\overline{x}) = \sum_{y \in \overline{x}} \pi^0(y) \forall \overline{x} \in \overline{S}_i$ holds in this case. Now:

$$\overline{\pi}_i^t(\overline{x}) = \sum_{k=0}^{\infty} e^{-\widehat{q}t} \frac{(\widehat{q}t)^k}{k!} \overline{\pi}_i^k(\overline{x}) \quad \text{(by statement 2(b) above)}$$

$$= \sum_{k=0}^{\infty} e^{-\widehat{q}t} \frac{(\widehat{q}t)^k}{k!} \sum_{y \in \overline{S}_i^{\overline{x}}} \pi^k(y) \quad \text{(using Theorem 3)}$$

$$= \sum_{y \in \overline{S}_i^{\overline{x}}} \sum_{k=0}^{\infty} e^{-\widehat{q}t} \frac{(\widehat{q}t)^k}{k!} \pi^k(y) \text{(using Tonelli's theorem [Tao, 2011])}$$

$$= \sum_{y \in \overline{S}_i^{\overline{x}}} \pi^t(y) \quad \text{(by statement 2(a) above).}$$

$\square$

Lemma 2. $R_t^{l,i} = \overline{R}_t^{l,i}$

Proof.

$$R_t^{l,i} = \sum_{x \in \mathcal{S}} (R^{l,i}(x))\pi^t(x) = \sum_{\overline{z} \in \overline{S}_i} \sum_{x \in \overline{S}_i^{\overline{z}}} (R^{l,i}(x))\pi^t(x)$$

$$\text{(Note: } R^{l,i}(x) = R^{l,i}(y) = \overline{R}^{l,i}(\overline{z}) \forall x, y \in \overline{S}_i^{\overline{z}} \text{ using Equations 3, 7}$$

$$\text{and the lumping criteria: } x_i = y_i = \overline{z} \forall x, y \in \overline{S}_i^{\overline{z}})$$

$$= \sum_{\overline{z} \in \overline{S}_i} \sum_{x \in \overline{S}_i^{\overline{z}}} \overline{R}^{l,i}(\overline{z})\pi^t(x) = \sum_{\overline{z} \in \overline{S}_i} \overline{R}^{l,i}(\overline{z}) \sum_{x \in \overline{S}_i^{\overline{z}}} \pi^t(x)$$

$$= \sum_{\overline{z} \in \overline{S}_i} \overline{R}^{l,i}(\overline{z})\overline{\pi}_i^t(\overline{z}) \quad \text{(Using Lemma 1)}$$

$$= \overline{R}_t^{l,i}$$

$\square$

Lemma 3. $R_t^I = \sum_{1 \le i \le n} c_i \overline{R}_t^{I,i}$

Proof.

$$R_t^I = \sum_{x \in \mathcal{S}} (R^I(x)) \pi^t(x) = \sum_{x \in \mathcal{S}} \sum_{1 \le i \le n} c_i (x_i \delta^{\log \mu} - \beta) \pi^t(x)$$

$$= \sum_{1 \le i \le n} \sum_{x \in \mathcal{S}} c_i (x_i \delta^{\log \mu} - \beta) \pi^t(x) = \sum_{1 \le i \le n} \sum_{\overline{z} \in \overline{S}_i} \sum_{x \in \overline{S}_i^{\overline{z}}} c_i (x_i \delta^{\log \mu} - \beta) \pi^t(x)$$

$$= \sum_{1 \le i \le n} c_i \sum_{\overline{z} \in \overline{S}_i} \sum_{x \in \overline{S}_i^{\overline{z}}} (x_i \delta^{\log \mu} - \beta) \pi^t(x)$$

$$(\text{Note: } \forall x, y \in \overline{S}_i^{\overline{z}}, (x_i \delta^{\log \mu} - \beta) = (y_i \delta^{\log \mu} - \beta) = \overline{R}^{I,i}(\overline{z})$$

$$\text{due to lumping critera } x_i = y_i = \overline{z} \forall x, y \in \overline{S}_i^{\overline{z}})$$

$$= \sum_{1 \le i \le n} c_i \sum_{\overline{z} \in \overline{S}_i} \sum_{x \in \overline{S}_i^{\overline{z}}} \overline{R}^{I,i}(\overline{z}) \pi^t(x) = \sum_{1 \le i \le n} c_i \sum_{\overline{z} \in \overline{S}_i} \overline{R}^{I,i}(\overline{z}) \sum_{x \in \overline{S}_i^{\overline{z}}} \pi^t(x)$$

$$= \sum_{1 \le i \le n} c_i \sum_{\overline{z} \in \overline{S}_i} \overline{R}^{I,i}(\overline{z}) \overline{\pi}_i^t(\overline{z}) \quad (\text{Using Lemma 1})$$

$$= \sum_{1 \le i \le n} c_i \overline{R}_t^{I,i}$$

$\square$

Lemma 4. *The CTMC $X(t) = (\mathcal{S}, Q)$ is irreducible and positive recurrent and a steady state vector Π exists for this CTMC.*

Proof. A steady state vector Π exists for a CTMC if it is irreducible and positive recurrent. And a CTMC $X(t)$ is irreducible and positive recurrent if the underlying DTMC Y_k is irreducible and positive recurrent. Also, a finite-DTMC is irreducible and positive recurrent if it has a single communicating class [Viswanadham and Narahari, 1992].

As argued in the proof of Lemma 1 an underlying DTMC $Y_k = (\mathcal{S}, P)$ exists for our CTMC $X(t) = (\mathcal{S}, Q)$ and the transition probability matrix P is given by: $P = Q/\widehat{q} + I$, where $\widehat{q}$ is an upper-bound on the entries in the Q matrix. Using Equation 1 we thus define P as:

$$P(x, y) = \begin{cases} \lambda_j / \widehat{q} & \text{if } \sum_i |x_i - y_i| = 1 \ \& \ \exists j : y_j - x_j = 1 \ (\text{Case } i) \\ \mu / \widehat{q} & \text{if } y = (0, \ldots, 0) \ne x \ (\text{Case } ii) \\ 1 + \sum_{y' \in \mathcal{S} \setminus \{x\}} \frac{-Q(x,y')}{\widehat{q}} & \text{if } x = y \ (\text{Case } iii) \\ 0 & o/w \ (\text{Case } iv). \end{cases} \tag{24}$$

We shall now prove that the DTMC Y_k has a single communicating class by arguing that the state $x : (0, 0, \ldots, 0)$ is reachable from each state $y \in \mathcal{S}$ and also each of these states y is reachable from the state x. This in turn implies that any two states y and z are accessible from each other.

- The one step transition probability $p_{yx}(1)$ from a state $y \in \mathcal{S} \setminus \{x\}$ to the state $x : (0, 0, \ldots, 0)$ is non-zero (i.e. Case (ii) in Equation 24). Also a state is always accessible from itself in 0 steps. Thus the state $x : (0, 0, \ldots, 0)$ is accessible from all states $y \in \mathcal{S}$.

- Consider a state $y = (y_1, y_2, \ldots, y_n) \in \mathcal{S}$. Now, if the DTMC Y_k is in state x and student 1 starts arriving on the forum one by one y_1 times then the following state sequence will be generated: $(0, 0, \ldots, 0), (1, 0, \ldots, 0), (2, 0, \ldots, 0), \ldots, (y_1, 0, \ldots, 0)$ and then student 2 starts arriving one by one y_2 times, then the following state sequence is incremented to the existing state sequence: $(y_1, 1, \ldots, 0), (y_1, 2, \ldots, 0), \ldots, (y_1, y_2, \ldots, 0)$. Similarly now if the student 3 arrives one by one y_3 times (then the DTMC Y_k reaches state $(y_1, y_2, y_3, 0, \ldots, 0)$) and so on till the student n also arrives one by one y_n times, we finally reach state $y : (y_1, y_2, \ldots, y_n)$.

    Now each of the transition between two consecutive states in this state sequence satisfies the Case (i) of Equation 24, thus all transitions in this sequence have some positive transition probability. Thus a $\sum_{i=1}^{n} y_i$ (number of transitions/steps) step walk exists from state x to state y and thus state y is accessible from state x [Gallager, 2012].

Thus all states in $\mathcal{S}$ are accessible from each other and so the DTMC Y_k has a single communicating class. Now as Y_k is finite and has a single communicating class, thus the underlying DTMC Y_k is irreducible and positive recurrent. Therefore the CTMC $X(t)$ is also irreducible and positive recurrent. Thus the steady state vector (or limiting probabilities) Π exist [Viswanadham and Narahari, 1992] for the CTMC $X(t)$. $\qquad\square$

Lemma 5. *The lumped-CTMC $\overline{X}_i(t) = (\overline{S}_i, \overline{Q}_i)$ is irreducible and positive recurrent and a steady state vector $\overline{\Pi}_i$ exists for this CTMC. Also each component of $\overline{\Pi}_i$ will be obtained as:* $\overline{\Pi}_i(\overline{x}) = \sum\limits_{x \in \overline{S}_i^{\overline{x}}} \Pi(x).$

Proof. The CTMC $X(t) = (\mathcal{S}, Q)$ is irreducible with stationary distribution Π (From Lemma 4). Also $\overline{X}_i(t) = (\overline{S}_i, \overline{Q}_i)$ is the lumped chain w.r.t. the partition $\overline{S}_i$ of the state space $\mathcal{S}$ of $X(t)$. Thus using Theorem 4 we can say that the lumped chain $\overline{X}_i(t)$ also has a stationary distribution $\overline{\Pi}_i$ and whose components will be obtained as: $\overline{\Pi}_i(\overline{x}) = \sum\limits_{x \in \overline{S}_i^{\overline{x}}} \Pi(x).$ $\quad\square$

Theorem 6. *The expected transient aggregate net-rewards over time T and the expected steady-state net-rewards received by the students and the instructor is the same when calculated using the original CTMC $X(t)$ or the n lumped CTMCs $\overline{X}_i(t)$ $i \in \{1, 2, \ldots, n\}$.*

- *Expected transient aggregate net-rewards over time T.*
  a) $R_T^{l,i} = \overline{R}_T^{l,i}$ b) $R_T^I = \overline{R}_T^I$.

- *Expected steady-state net-rewards.*
  a) $R_\Pi^{l,i} = \overline{R}_\Pi^{l,i}$ b) $R_\Pi^I = \overline{R}_\Pi^I$.

Proof.   1. Expected transient aggregate net-rewards over time T
   (a) To Student i.

$$R_T^{l,i} = \int\limits_{t=0}^{T} R_t^{l,i} dt = \int\limits_{t=0}^{T} \overline{R}_t^{l,i} dt \ \text{ (Using Lemma 2)} = \overline{R}_T^{l,i}$$

   (b) To Instructor.

$$R_T^I = \int\limits_{t=0}^{T} R_t^I \ dt = \int\limits_{t=0}^{T} \sum_{1 \leq i \leq n} c_i \overline{R}_t^{I,i} \ dt \ \text{ (Using Lemma 3)}$$

$$= \sum_{1 \leq i \leq n} c_i \int\limits_{t=0}^{T} \overline{R}_t^{I,i} \ dt = \sum_{1 \leq i \leq n} c_i \overline{R}_T^{I,i} \ dt = \overline{R}_T^I$$

   2. Expected steady-state net-rewards
   (a) To Student i.

$$R_\Pi^{l,i} = \sum_{x \in \mathcal{S}} (R^{l,i}(x)) \Pi(x) = \sum_{\overline{z} \in \overline{S}_i} \sum_{x \in \overline{S}_i^{\overline{z}}} (R^{l,i}(x)) \Pi(x)$$

$$\text{(Note: } R^{l,i}(x) = R^{l,i}(y) = \overline{R}^{l,i}(\overline{z}) \forall x, y \in \overline{S}_i^{\overline{z}} \text{ using Equations 3, 7}$$

$$\text{and the lumping criteria: } x_i = y_i = \overline{z} \forall x, y \in \overline{S}_i^{\overline{z}})$$

$$= \sum_{\overline{z} \in \overline{S}_i} \sum_{x \in \overline{S}_i^{\overline{z}}} \overline{R}^{l,i}(\overline{z}) \Pi(x) = \sum_{\overline{z} \in \overline{S}_i} \overline{R}^{l,i}(\overline{z}) \sum_{x \in \overline{S}_i^{\overline{z}}} \Pi(x)$$

$$= \sum_{\overline{z} \in \overline{S}_i} \overline{R}^{l,i}(\overline{z}) \overline{\Pi}_i(\overline{z}) \quad \text{(Using Lemma 5)}$$

$$= \overline{R}_\Pi^{l,i}$$

(b) To Instructor.

$$R_{\Pi}^I = \sum_{x \in \mathcal{S}} R^I(x)\Pi(x) = \sum_{x \in \mathcal{S}} \sum_{1 \leq i \leq n} c_i(x_i \delta^{\log \mu} - \beta)\Pi(x) = \sum_{1 \leq i \leq n} \sum_{x \in \mathcal{S}} c_i(x_i \delta^{\log \mu} - \beta)\Pi(x)$$

$$= \sum_{1 \leq i \leq n} \sum_{\overline{z} \in \overline{S}_i} \sum_{x \in \overline{S}_i^{\overline{z}}} c_i(x_i \delta^{\log \mu} - \beta)\Pi(x) = \sum_{1 \leq i \leq n} c_i \sum_{\overline{z} \in \overline{S}_i} \sum_{x \in \overline{S}_i^{\overline{z}}} (x_i \delta^{\log \mu} - \beta)\Pi(x)$$

(Note: $\forall x, y \in \overline{S}_i^{\overline{z}}, (x_i \delta^{\log \mu} - \beta) = (y_i \delta^{\log \mu} - \beta) = \overline{R}^{I,i}(\overline{z})$)

$$= \sum_{1 \leq i \leq n} c_i \sum_{\overline{z} \in \overline{S}_i} \sum_{x \in \overline{S}_i^{\overline{z}}} \overline{R}^{I,i}(\overline{z})\Pi(x) = \sum_{1 \leq i \leq n} c_i \sum_{\overline{z} \in \overline{S}_i} \overline{R}^{I,i}(\overline{z}) \sum_{x \in \overline{S}_i^{\overline{z}}} \Pi(x)$$

$$= \sum_{1 \leq i \leq n} c_i \sum_{\overline{z} \in \overline{S}_i} \overline{R}^{I,i}(\overline{z})\overline{\Pi}_i(\overline{z}) \quad \text{(Using Lemma 5)}$$

$$= \sum_{1 \leq i \leq n} c_i \overline{R}_{\Pi}^{I,i} = \overline{R}_{\Pi}^I$$

$\square$

B　Proofs : Steady State Analysis

Lemma 6. *The solution to equations 9, 10, and 11 is:*

$$\overline{\Pi}_i(\overline{x}) = \frac{\rho_i^{\overline{x}}}{(1 + \rho_i)^{\overline{x}+1}}, \forall \overline{x} : 0 \leq \overline{x} < M \text{ and } \quad \overline{\Pi}_i(M) = \frac{\rho_i^M}{(1 + \rho_i)^M}. \text{ where, } \rho_i = \frac{\lambda_i}{\mu}.$$

Proof.

Inflow in state 0 = Outflow from state 0

$$\sum_{\overline{y} \geq 1} \mu \overline{\Pi}_i(\overline{y}) = \lambda_i \overline{\Pi}_i(0)$$

$$\Rightarrow \mu(1 - \overline{\Pi}_i(0)) = \lambda_i \overline{\Pi}_i(0) \Rightarrow \overline{\Pi}_i(0)(\mu + \lambda_i) = \mu$$

$$\Rightarrow \overline{\Pi}_i(0) = \frac{\mu}{\mu + \lambda_i} = \frac{1}{1 + \rho_i} \text{ (Taking } \rho_i = \frac{\lambda_i}{\mu}\text{)}$$

Similarly for any state $\overline{x} \in \{1, \ldots, M - 1\}$

Inflow in state $\overline{x}$ = Outflow from state $\overline{x}$

$$\lambda_i \overline{\Pi}_i(\overline{x} - 1) = \mu \overline{\Pi}_i(\overline{x}) + \lambda_i \overline{\Pi}_i(\overline{x})$$

$$\Rightarrow \overline{\Pi}_i(\overline{x}) = \frac{\lambda_i}{\mu + \lambda_i} \overline{\Pi}_i(\overline{x} - 1) = \frac{\rho_i}{1 + \rho_i} \overline{\Pi}_i(\overline{x} - 1)$$

Expanding RHS till $\overline{\Pi}_i(0)$, we get

$$\Rightarrow \overline{\Pi}_i(\overline{x}) = \left(\frac{\rho_i}{1 + \rho_i}\right)^{\overline{x}} \overline{\Pi}_i(0) = \frac{\rho_i^{\overline{x}}}{(1 + \rho_i)^{\overline{x}+1}}$$

Inflow in state M = Outflow from state M

$$\lambda_i \overline{\Pi}_i(M - 1) = \mu \overline{\Pi}_i(M)$$

$$\Rightarrow \overline{\Pi}_i(M) = \frac{\lambda_i}{\mu} \overline{\Pi}_i(M - 1)$$

$$\Rightarrow \overline{\Pi}_i(M) = \rho_i \frac{\rho_i^{M-1}}{(1 + \rho_i)^M} = \frac{\rho_i^M}{(1 + \rho_i)^M}$$

$\square$

Claim 1.

$$\sum_{\overline{x}=0}^{m_l} \overline{x}\overline{\Pi}_i(\overline{x}) = \rho_i \left(1 - (m_l + 1) \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l} + m_l \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l+1} \right)$$

Proof.

$$\sum_{\overline{x}=0}^{m_l} \overline{x}\overline{\Pi}_i(\overline{x}) = \sum_{\overline{x}=0}^{m_l} \overline{x}\frac{\rho_i^{\overline{x}}}{(1 + \rho_i)^{\overline{x}+1}} = \frac{1}{(1 + \rho_i)} \sum_{\overline{x}=0}^{m_l} \overline{x} \left(\frac{\rho_i}{1 + \rho_i} \right)^{\overline{x}}$$

$$= \frac{1 - (m_l + 1) \left(\frac{\rho_i}{1+\rho_i} \right)^{m_l} + m_l \left(\frac{\rho_i}{1+\rho_i} \right)^{m_l+1}}{\left(\frac{1}{1+\rho_i} \right)^2} \times \frac{1}{(1 + \rho_i)} \frac{\rho_i}{(1 + \rho_i)}$$

$$= \rho_i \left(1 - (m_l + 1) \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l} + m_l \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l+1} \right)$$

$\square$

Claim 2. $\sum_{\overline{x}=0}^{m_l} \overline{\Pi}_i(\overline{x}) = 1 - \left(\frac{\rho_i}{1+\rho_i} \right)^{m_l+1}.$

Proof.

$$\sum_{\overline{x}=0}^{m_l} \overline{\Pi}_i(\overline{x}) = \sum_{\overline{x}=0}^{m_l} \frac{\rho_i^{\overline{x}}}{(1 + \rho_i)^{\overline{x}+1}} = \frac{1}{(1 + \rho_i)} \sum_{\overline{x}=0}^{m_l} \left(\frac{\rho_i}{1 + \rho_i} \right)^{\overline{x}} = \frac{1}{(1 + \rho_i)} \frac{1 - \left(\frac{\rho_i}{1+\rho_i} \right)^{m_l+1}}{1 - \frac{\rho_i}{1+\rho_i}}$$

$$= 1 - \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l+1}$$

$\square$

Claim 3. $\sum_{\overline{x}=0}^{M} \overline{x}\overline{\Pi}_i(\overline{x}) = \rho_i \left(1 - \left(\frac{\rho_i}{1+\rho_i} \right)^{M} \right).$

Proof.

$$\sum_{\overline{x}=0}^{M} \overline{x}\overline{\Pi}_i(\overline{x}) = \sum_{\overline{x}=0}^{M-1} \overline{x}\overline{\Pi}_i(\overline{x}) + M\overline{\Pi}_i(M)$$

(Using Claim 1)

$$= \rho_i \left(1 - M \left(\frac{\rho_i}{1 + \rho_i} \right)^{M-1} + (M - 1) \left(\frac{\rho_i}{1 + \rho_i} \right)^{M} \right) + M \left(\frac{\rho_i}{1 + \rho_i} \right)^{M}$$

$$= \rho_i \left(1 - M \left(\frac{\rho_i}{1 + \rho_i} \right)^{M-1} - \left(\frac{\rho_i}{1 + \rho_i} \right)^{M} \right) + M (1 + \rho_i) \left(\frac{\rho_i}{1 + \rho_i} \right)^{M}$$

$$= \rho_i \left(1 - M \left(\frac{\rho_i}{1 + \rho_i} \right)^{M-1} - \left(\frac{\rho_i}{1 + \rho_i} \right)^{M} \right) + M \left(\frac{\rho_i}{1 + \rho_i} \right)^{M-1}$$

$$= \rho_i \left(1 - \left(\frac{\rho_i}{1 + \rho_i} \right)^{M} \right)$$

$\square$

Theorem 7. *The expected steady state net-rewards of the student i and the instructor in $\overline{X}_i(t)$ are given by:*

$$(a)\ \overline{R}_\Pi^{l,i} = \delta^{\log \mu} \rho_i \left(1 - \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l} \right) - \alpha_l \rho_i \left(1 - \left(\frac{\rho_i}{1 + \rho_i} \right)^{M} \right)$$

$$(b)\ \overline{R}_\Pi^{I,i} = \delta^{\log \mu} \rho_i \left(1 - \left(\frac{\rho_i}{1 + \rho_i} \right)^{M} \right) - \beta$$

Proof. (a)

$$\overline{R}_\Pi^{l,i} = \sum_{\overline{x} \in \overline{S}_i} \overline{R}^{l,i}(\overline{x}) \overline{\Pi}_i(\overline{x}) = \sum_{\overline{x}=0}^{M} \overline{R}^{l,i}(\overline{x}) \overline{\Pi}_i(\overline{x}) = \sum_{\overline{x}=0}^{M} \left(\overline{r}^{l,i}(\overline{x}) - \alpha_l \overline{x} \right) \overline{\Pi}_i(\overline{x})$$

$$= \sum_{\overline{x}=0}^{m_l} \left(\overline{x} \delta^{\log \mu} - \alpha_l \overline{x} \right) \overline{\Pi}_i(\overline{x}) + \sum_{\overline{x}=m_l+1}^{M} \left(m_l \delta^{\log \mu} - \alpha_l \overline{x} \right) \overline{\Pi}_i(\overline{x})$$

$$= \sum_{\overline{x}=0}^{m_l} \left(\overline{x} \delta^{\log \mu} \right) \overline{\Pi}_i(\overline{x}) + \sum_{\overline{x}=m_l+1}^{M} \left(m_l \delta^{\log \mu} \right) \overline{\Pi}_i(\overline{x}) - \sum_{\overline{x}=0}^{M} \alpha_l \overline{x} \overline{\Pi}_i(\overline{x})$$

$$= \sum_{\overline{x}=0}^{m_l} \left(\overline{x} \delta^{\log \mu} \right) \overline{\Pi}_i(\overline{x}) + \left(m_l \delta^{\log \mu} \right) \sum_{\overline{x}=m_l+1}^{M} \overline{\Pi}_i(\overline{x}) - \sum_{\overline{x}=0}^{M} \alpha_l \overline{x} \overline{\Pi}_i(\overline{x})$$

$$= \delta^{\log \mu} \sum_{\overline{x}=0}^{m_l} \overline{x} \overline{\Pi}_i(\overline{x}) + \left(m_l \delta^{\log \mu} \right) \left(1 - \sum_{\overline{x}=0}^{m_l} \overline{\Pi}_i(\overline{x}) \right) - \sum_{\overline{x}=0}^{M} \alpha_l \overline{x} \overline{\Pi}_i(\overline{x})$$

(Using Claims 1 and 2)

$$= \delta^{\log \mu} \rho_i \left(1 - (m_l + 1) \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l} + m_l \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l+1} \right) - \sum_{\overline{x}=0}^{M} \alpha_l \overline{x} \overline{\Pi}_i(\overline{x}) + \left(m_l \delta^{\log \mu} \right) \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l+1}$$

$$= \delta^{\log \mu} \rho_i \left(1 - (m_l + 1) \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l} \right) + \left(m_l \delta^{\log \mu} \right) \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l+1} (1 + \rho_i) - \sum_{\overline{x}=0}^{M} \alpha_l \overline{x} \overline{\Pi}_i(\overline{x})$$

$$= \delta^{\log \mu} \rho_i \left(1 - (m_l + 1) \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l} \right) + \left(m_l \delta^{\log \mu} \right) \frac{\rho_i^{m_l+1}}{(1 + \rho_i)^{m_l}} - \sum_{\overline{x}=0}^{M} \alpha_l \overline{x} \overline{\Pi}_i(\overline{x})$$

$$= \delta^{\log \mu} \rho_i \left(1 - \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l} \right) - \sum_{\overline{x}=0}^{M} \alpha_l \overline{x} \overline{\Pi}_i(\overline{x})$$

$$= \delta^{\log \mu} \rho_i \left(1 - \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l} \right) - \alpha_l \sum_{\overline{x}=0}^{M} \overline{x} \overline{\Pi}_i(\overline{x})$$

(Using Claim 3)

$$= \delta^{\log \mu} \rho_i \left(1 - \left(\frac{\rho_i}{1 + \rho_i} \right)^{m_l} \right) - \alpha_l \rho_i \left(1 - \left(\frac{\rho_i}{1 + \rho_i} \right)^{M} \right)$$

$\square$

Proof. (b)

$$\overline{R}_{\Pi}^{I,i} = \sum_{\overline{x}=0}^{M} \left(\overline{R}^{I,i}(\overline{x}) \right) \overline{\Pi}_i(\overline{x}) = \sum_{\overline{x}=0}^{M} \left(\overline{r}^{I,i}(\overline{x}) - \beta \right) \overline{\Pi}_i(\overline{x}) = \sum_{\overline{x}=0}^{M} \left(\overline{x} \delta^{\log \mu} - \beta \right) \overline{\Pi}_i(\overline{x})$$

$$= \sum_{\overline{x}=0}^{M} \overline{x} \delta^{\log \mu} \overline{\Pi}_i(\overline{x}) - \sum_{\overline{x}=0}^{M} \beta \overline{\Pi}_i(\overline{x}) = \delta^{\log \mu} \sum_{\overline{x}=0}^{M} \overline{x} \overline{\Pi}_i(\overline{x}) - \beta$$

$$= \delta^{\log \mu} \rho_i \left(1 - \left(\frac{\rho_i}{1 + \rho_i} \right)^M \right) - \beta \quad \text{(Using Claim 3)}$$

$\square$

C   Proofs : Transient Analysis

Lemma 7. *Given the initial distribution $\overline{\pi}_i^0$ for $\overline{X}_i(t)$ as $\overline{\pi}_i^0(0) = 1$ and $\overline{\pi}_i^0(\overline{x}) = 0 \; \forall \overline{x} \in \{1, 2, \ldots, M\}$. The solution to the differential equations 12, 13, and 14 is given by:*

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

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

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

Proof.   • **For state** 0

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

$$\Rightarrow \frac{d\overline{\pi}_i^t(0)}{dt} + \lambda_i \overline{\pi}_i^t(0) = \mu(1 - \overline{\pi}_i^t(0))$$

$$\Rightarrow \frac{d\overline{\pi}_i^t(0)}{dt} + (\lambda_i + \mu)\overline{\pi}_i^t(0) = \mu$$

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

(Now using the initial distributions $\overline{\pi}_i^0(0) = 1$)

$$\overline{\pi}_i^0(0) = \frac{\mu}{\mu + \lambda_i} + ce^{-(\mu+\lambda_i)0} = 1 \Rightarrow c = 1 - \frac{\mu}{\mu + \lambda_i} = \frac{\lambda_i}{\mu + \lambda_i}$$

Thus we get, $\overline{\pi}_i^t(0) = \frac{\mu}{K_i} + \frac{\lambda_i}{K_i} e^{-K_i t}$, where $K_i = \mu + \lambda_i$.

• **For state** 1

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

$$\Rightarrow \frac{d\overline{\pi}_i^t(1)}{dt} = -(\lambda_i + \mu)\overline{\pi}_i^t(1) + \lambda_i \left(\frac{\mu}{K_i} + \frac{\lambda_i}{K_i} e^{-K_i t} \right)$$

$$\Rightarrow \frac{d\overline{\pi}_i^t(1)}{dt} + K_i \overline{\pi}_i^t(1) = W$$

where $K_i = (\mu + \lambda_i)$ and $W = \left(\frac{\lambda_i}{K_i} \left(\mu + \lambda_i e^{-K_i t} \right) \right)$

$$\Rightarrow \overline{\pi}_i^t(1) e^{\int K_i dt} = c + \int W e^{\int K_i dt} dt$$

$$\Rightarrow \overline{\pi}_i^t(1) e^{K_i t} = c + \int \left(\frac{\lambda_i}{K_i} \left(\mu + \lambda_i e^{-K_i t} \right) \right) e^{K_i t} dt$$

$$\Rightarrow \overline{\pi}_i^t(1) e^{K_i t} = c + \frac{\lambda_i \mu}{K_i} \int e^{K_i t} dt + \frac{\lambda_i^2}{K_i} \int 1 dt$$

$$\Rightarrow \overline{\pi}_i^t(1) e^{K_i t} = c + \frac{\lambda_i \mu}{K_i} \frac{e^{K_i t}}{K_i} + \frac{\lambda_i^2}{K_i} t$$

$$\Rightarrow \overline{\pi}_i^t(1) = \frac{\lambda_i \mu}{K_i^2} + \left(c + \frac{\lambda_i^2}{K_i} t \right) e^{-K_i t}$$

(Now using the initial distribution $\overline{\pi}_i^0(1) = 0$)

$$\overline{\pi}_i^0(1) = \frac{\lambda_i \mu}{K_i^2} + (c) e^0 = 0 \Rightarrow c = -\frac{\lambda_i \mu}{K_i^2}$$

Thus we get, $\overline{\pi}_i^t(1) = \dfrac{\lambda_i \mu}{K_i^2} + \left(-\dfrac{\lambda_i \mu}{K_i^2} + \dfrac{\lambda_i^2 t}{K_i} \right) e^{-K_i t}$

- Similarly on solving for **States 2 and 3** we get:

$$\overline{\pi}_i^t(2) = \frac{\lambda_i^2 \mu}{K_i^3} + \left(-\frac{\lambda_i^2 \mu}{K_i^3} + \frac{\lambda_i^3 t^2}{2 K_i} - \frac{\lambda_i^2 \mu t}{K_i^2} \right) e^{-K_i t}$$

$$\overline{\pi}_i^t(3) = \frac{\lambda_i^3 \mu}{K_i^4} + \left(-\frac{\lambda_i^3 \mu}{K_i^4} - \frac{\lambda_i^3 \mu t}{K_i^3} + \frac{\lambda_i^4 t^3}{3! K_i} - \frac{\lambda_i^3 \mu t^2}{2 K_i^2} \right) e^{-K_i t}$$

where $K_i = (\mu + \lambda_i)$

- Observing from $\overline{\pi}_i^t(1), \overline{\pi}_i^t(2), \overline{\pi}_i^t(3)$, we claim and prove by induction that $\forall \overline{x} \in \{1, \ldots, M-1\}$:

$$\overline{\pi}_i^t(\overline{x}) = \left(-\left(\frac{\lambda_i}{K_i} \right)^{\overline{x}} \frac{\mu}{K_i} - \sum_{\overline{y}=1}^{\overline{x}-1} \frac{\lambda_i^{\overline{x}} t^{\overline{y}} \mu}{\overline{y}! K_i^{\overline{x}-\overline{y}+1}} + \frac{\lambda_i^{\overline{x}+1} t^{\overline{x}}}{\overline{x}! K_i} \right) e^{-K_i t} + \left(\frac{\lambda_i}{K_i} \right)^{\overline{x}} \frac{\mu}{K_i} \quad (\text{ where } K_i = (\mu + \lambda_i)) \tag{25}$$

Now let us assume that this holds for some state $\overline{z}$ s.t. $1 \leq \overline{z} < M-1$. Thus we have the value of $\overline{\pi}_i^t(\overline{z})$ by substituting in place of $\overline{x}$ in the equation above. Now we shall prove that the Equation 25 holds for state $\overline{z}+1$ also.

For state $\overline{z}+1$

$$\frac{d\overline{\pi}_i^t(\overline{z}+1)}{dt} = -(\lambda_i + \mu)\overline{\pi}_i^t(\overline{z}+1) + \lambda_i \overline{\pi}_i^t(\overline{z})$$

$$\Rightarrow \frac{d\overline{\pi}_i^t(\overline{z}+1)}{dt} + (\lambda_i + \mu)\overline{\pi}_i^t(\overline{z}+1) = \lambda_i \overline{\pi}_i^t(\overline{z})$$

$$\Rightarrow \overline{\pi}_i^t(\overline{z}+1)e^{K_i t} = c + \int W e^{K_i t} dt$$

(where $K_i = (\lambda_i + \mu)$ and $W = \lambda_i \overline{\pi}_i^t(\overline{z})$)

$$\Rightarrow \overline{\pi}_i^t(\overline{z}+1)e^{K_i t} = c + \int \left(\frac{\lambda_i}{K_i}\right)^{\overline{z}} \frac{\lambda_i \mu}{K_i} e^{K_i t} dt - \int \left(\frac{\lambda_i}{K_i}\right)^{\overline{z}} \frac{\lambda_i \mu}{K_i} dt - \sum_{\overline{y}=1}^{\overline{z}-1} \int \frac{\lambda^{\overline{z}+1} \mu t^{\overline{y}}}{\overline{y}! K_i^{\overline{z}-\overline{y}+1}} dt + \int \frac{\lambda_i^{\overline{z}+2} t^{\overline{z}}}{\overline{z}! K_i} dt$$

$$\Rightarrow \overline{\pi}_i^t(\overline{z}+1)e^{K_i t} = c + \left(\frac{\lambda_i}{K_i}\right)^{\overline{z}+1} \frac{\mu}{K_i} e^{K_i t} - \left(\frac{\lambda_i}{K_i}\right)^{\overline{z}+1} \mu t - \sum_{\overline{y}=1}^{\overline{z}-1} \frac{\lambda^{\overline{z}+1} \mu t^{\overline{y}+1}}{(\overline{y}+1)! K_i^{\overline{z}-\overline{y}+1}} + \frac{\lambda_i^{\overline{z}+2} t^{\overline{z}+1}}{(\overline{z}+1)! K_i}$$

Now, set $\overline{u} = \overline{y} + 1$ in the summation.

$$\Rightarrow \overline{\pi}_i^t(\overline{z}+1)e^{K_i t} = c + \left(\frac{\lambda_i}{K_i}\right)^{\overline{z}+1} \frac{\mu}{K_i} e^{K_i t} - \left(\frac{\lambda_i}{K_i}\right)^{\overline{z}+1} \mu t - \sum_{\overline{u}=2}^{\overline{z}} \frac{\lambda^{\overline{z}+1} \mu t^{\overline{u}}}{(\overline{u})! K_i^{\overline{z}-\overline{u}+2}} + \frac{\lambda_i^{\overline{z}+2} t^{\overline{z}+1}}{(\overline{z}+1)! K_i}$$

Now using the initial distribution $\overline{\pi}_i^0(\overline{z}+1) = 0$ we get $c = -\left(\frac{\lambda_i}{K_i}\right)^{\overline{z}+1} \frac{\mu}{K_i}$ and putting this

value above we get,

$$\Rightarrow \overline{\pi}_i^t(\overline{z}+1) = \left(\frac{\lambda_i}{K_i}\right)^{\overline{z}+1} \frac{\mu}{K_i} + \left(-\left(\frac{\lambda_i}{K_i}\right)^{\overline{z}+1} \mu t - \sum_{\overline{u}=2}^{\overline{z}} \frac{\lambda^{\overline{z}+1} \mu t^{\overline{u}}}{(\overline{u})! K_i^{\overline{z}-\overline{u}+2}} + \frac{\lambda_i^{\overline{z}+2} t^{\overline{z}+1}}{(\overline{z}+1)! K_i}\right) e^{-K_i t}$$

$$- \left(\left(\frac{\lambda_i}{K_i}\right)^{\overline{z}+1} \frac{\mu}{K_i}\right) e^{-K_i t}$$

$$\Rightarrow \overline{\pi}_i^t(\overline{z}+1) = \left(\frac{\lambda_i}{K_i}\right)^{\overline{z}+1} \frac{\mu}{K_i} + \left(-\left(\frac{\lambda_i}{K_i}\right)^{\overline{z}+1} \frac{\mu}{K_i} - \sum_{\overline{u}=1}^{\overline{z}} \frac{\lambda^{\overline{z}+1} \mu t^{\overline{u}}}{(\overline{u})! K_i^{\overline{z}-\overline{u}+2}} + \frac{\lambda_i^{\overline{z}+2} t^{\overline{z}+1}}{(\overline{z}+1)! K_i}\right) e^{-K_i t}$$

which is what we will get if we replace $\overline{x}$ by $(\overline{z}+1)$ in Equation 25.

Hence proved by induction that we may calculate the transient distributions of states between 1 and $M-1$ using Equation 25.

- Now as the total probability of being in any of the states at any time t is equal to 1, we calculate the transient probability $\overline{\pi}_i^t(M)$ of being in state M at time t as:

$$\overline{\pi}_i^t(M) = 1 - \sum_{\overline{x}=0}^{M-1} \overline{\pi}_i^t(\overline{x})$$

$\square$

Theorem 8. *The expected transient aggregate net-rewards of the student i and the instructor in $\overline{X}_i(t)$ are given by:*

$$(a)\ \overline{R}_T^{l,i} = \delta^{\log \mu} \sum_{\overline{x}=0}^{m_l} (\overline{x} - m_l) \int_0^T \overline{\pi}_i^t(\overline{x}) dt + m_l \delta^{\log \mu} T - \alpha_l \sum_{\overline{x}=0}^M \overline{x} \int_0^T \overline{\pi}_i^t(\overline{x}) dt$$

$$(b)\ \overline{R}_T^{I,i} = \delta^{\log \mu} \sum_{\overline{x}=0}^M \overline{x} \int_0^T \overline{\pi}_i^t(\overline{x}) dt - \beta T.$$

Proof. (i)

$$\overline{R}_T^{l,i} = \int_0^T \sum_{\overline{x} \in \overline{S}_i} \overline{R}^{l,i}(\overline{x}) \overline{\pi}_i^t(\overline{x}) dt = \int_0^T \sum_{\overline{x}=0}^M \overline{R}^{l,i}(\overline{x}) \overline{\pi}_i^t(\overline{x}) dt$$

$$= \int_0^T \sum_{\overline{x}=0}^M \left(\overline{r}^{l,i}(\overline{x}) - \alpha_l \overline{x} \right) \overline{\pi}_i^t(\overline{x}) dt$$

$$= \int_0^T \sum_{\overline{x}=0}^{m_l} \left(\overline{x} \delta^{\log \mu} - \alpha_l \overline{x} \right) \overline{\pi}_i^t(\overline{x}) dt + \int_0^T \sum_{\overline{x}=m_l+1}^M \left(m_l \delta^{\log \mu} - \alpha_l \overline{x} \right) \overline{\pi}_i^t(\overline{x}) dt$$

$$= \int_0^T \sum_{\overline{x}=0}^{m_l} \overline{x} \delta^{\log \mu} \overline{\pi}_i^t(\overline{x}) dt - \int_0^T \sum_{\overline{x}=0}^M \alpha_l \overline{x} \overline{\pi}_i^t(\overline{x}) dt + \int_0^T \sum_{\overline{x}=m_l+1}^M m_l \delta^{\log \mu} \overline{\pi}_i^t(\overline{x}) dt$$

$$= \delta^{\log \mu} \int_0^T \sum_{\overline{x}=0}^{m_l} \overline{x} \overline{\pi}_i^t(\overline{x}) dt - \alpha_l \int_0^T \sum_{\overline{x}=0}^M \overline{x} \overline{\pi}_i^t(\overline{x}) dt + m_l \delta^{\log \mu} \int_0^T \sum_{\overline{x}=m_l+1}^M \overline{\pi}_i^t(\overline{x}) dt$$

$$= \delta^{\log \mu} \int_0^T \sum_{\overline{x}=0}^{m_l} \overline{x} \overline{\pi}_i^t(\overline{x}) dt - \alpha_l \int_0^T \sum_{\overline{x}=0}^M \overline{x} \overline{\pi}_i^t(\overline{x}) dt + m_l \delta^{\log \mu} \int_0^T \left(1 - \sum_{\overline{x}=0}^{m_l} \overline{\pi}_i^t(\overline{x}) \right) dt$$

$$= \delta^{\log \mu} \int_0^T \sum_{\overline{x}=0}^{m_l} \overline{x} \overline{\pi}_i^t(\overline{x}) dt - \alpha_l \int_0^T \sum_{\overline{x}=0}^M \overline{x} \overline{\pi}_i^t(\overline{x}) dt + m_l \delta^{\log \mu} T - m_l \delta^{\log \mu} \int_0^T \sum_{\overline{x}=0}^{m_l} \overline{\pi}_i^t(\overline{x}) dt$$

$$= m_l \delta^{\log \mu} T + \delta^{\log \mu} \int_0^T \sum_{\overline{x}=0}^{m_l} (\overline{x} - m_l) \overline{\pi}_i^t(\overline{x}) dt - \alpha_l \int_0^T \sum_{\overline{x}=0}^M \overline{x} \overline{\pi}_i^t(\overline{x}) dt$$

Interchanging the integral and the summation we get:

$$= m_l \delta^{\log \mu} T + \delta^{\log \mu} \sum_{\overline{x}=0}^{m_l} \int_0^T (\overline{x} - m_l) \overline{\pi}_i^t(\overline{x}) dt - \alpha_l \sum_{\overline{x}=0}^M \int_0^T \overline{x} \overline{\pi}_i^t(\overline{x}) dt$$

$$= m_l \delta^{\log \mu} T + \delta^{\log \mu} \sum_{\overline{x}=0}^{m_l} (\overline{x} - m_l) \int_0^T \overline{\pi}_i^t(\overline{x}) dt - \alpha_l \sum_{\overline{x}=0}^M \overline{x} \int_0^T \overline{\pi}_i^t(\overline{x}) dt$$

□

Proof. (ii)

$$\overline{R}_T^{I,i} = \int\limits_0^T \sum_{\overline{x}=0}^M \left(\overline{R}^{I,i}(\overline{x}) \right) \overline{\pi}_i^t(\overline{x}) dt$$

$$= \int\limits_0^T \sum_{\overline{x}=0}^M \left(\overline{r}^{I,i}(\overline{x}) - \beta \right) \overline{\pi}_i^t(\overline{x}) dt = \int\limits_0^T \sum_{\overline{x}=0}^M \left(\overline{x}\delta^{\log \mu} - \beta \right) \overline{\pi}_i^t(\overline{x}) dt$$

$$= \int\limits_0^T \sum_{\overline{x}=0}^M \overline{x}\delta^{\log \mu} \overline{\pi}_i^t(\overline{x}) dt - \int\limits_0^T \sum_{\overline{x}=0}^M \beta \overline{\pi}_i^t(\overline{x}) dt$$

$$= \delta^{\log \mu} \int\limits_0^T \sum_{\overline{x}=0}^M \overline{x}\overline{\pi}_i^t(\overline{x}) dt - \int\limits_0^T \beta dt = \delta^{\log \mu} \int\limits_0^T \sum_{\overline{x}=0}^M \overline{x}\overline{\pi}_i^t(\overline{x}) dt - \beta T$$

Interchanging the integral and the summation we get:

$$= \delta^{\log \mu} \sum_{\overline{x}=0}^M \overline{x} \int\limits_0^T \overline{\pi}_i^t(\overline{x}) dt - \beta T$$

$\square$

Claim 4.

$$\int\limits_0^T \overline{\pi}_i^t(\overline{x}) \, dt = \begin{cases} \frac{\mu T}{K_i} + \frac{\lambda_i(-e^{-K_iT}+1)}{K_i^2} & \text{if } \overline{x} = 0 \\ = \left(\frac{\lambda_i}{K_i} \right)^{\overline{x}} \frac{\mu}{K_i} T + \frac{\lambda_i^{\overline{x}}\mu}{K_l^{\overline{x}+2}} \left(e^{-K_iT} - \overline{x} \right) \\ + \frac{\lambda_i^{\overline{x}}\mu}{K_l^{\overline{x}+2}} \sum_{\overline{y}=1}^{\overline{x}-1} \left(\frac{\Gamma(\overline{y}+1,K_iT)}{\overline{y}!} \right) + \frac{\lambda_i^{\overline{x}+1}}{K_i^{\overline{x}+2}} \left(1 - \frac{\Gamma(\overline{x}+1,K_iT)}{\overline{x}!} \right) & \text{if } 0 < \overline{x} < M \\ = T - \sum_{\overline{x}=0}^{M-1} \int\limits_0^T \overline{\pi}_i^t(\overline{x}) dt & \text{if } \overline{x} = M. \end{cases}$$

$$\text{where } K_i = \lambda_i + \mu.$$

Proof. **State** 0

$$\int\limits_0^T \overline{\pi}_i^t(\overline{0}) \, dt = \int\limits_0^T \left(\frac{\mu}{K_i} + \frac{\lambda_i}{K_i} e^{-K_it} \right) \, dt$$

$$= \frac{\mu}{K_i} T + \left[\frac{\lambda_i}{K_i} \frac{e^{-K_it}}{-K_i} \right]_0^T = \frac{\mu}{K_i} T + \frac{\lambda_i}{(K_i)^2} \left(1 - e^{-K_iT} \right)$$

42

State $x \; \forall \overline{x} : 1 \leq \overline{x} \leq M - 1$

$$\int_0^T \overline{\pi}_i^t(\overline{x}) \, dt$$

$$= \int_0^T \left(- \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{K_i} - \sum_{\overline{y}=1}^{\overline{x}-1} \frac{\lambda_i^{\overline{x}} t^{\overline{y}} \mu}{\overline{y}! K_i^{\overline{x}-\overline{y}+1}} + \frac{\lambda_i^{\overline{x}+1} t^{\overline{x}}}{\overline{x}! K_i} \right) e^{-K_i t} dt + \int_0^T \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{K_i} \, dt$$

$$= \int_0^T \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{K_i} \, dt - \int_0^T \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{K_i} e^{-K_i t} dt - \int_0^T \sum_{\overline{y}=1}^{\overline{x}-1} \frac{\lambda_i^{\overline{x}} t^{\overline{y}} \mu}{\overline{y}! K_i^{\overline{x}-\overline{y}+1}} e^{-K_i t} \, dt + \int_0^T \frac{\lambda_i^{\overline{x}+1} t^{\overline{x}}}{\overline{x}! K_i} e^{-K_i t} \, dt$$

$$= \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{K_i} T - \left[\left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{K_i} \frac{e^{-K_i t}}{-K_i} \right]_0^T - \sum_{\overline{y}=1}^{\overline{x}-1} \frac{\lambda_i^{\overline{x}} \mu}{\overline{y}! K_i^{\overline{x}-\overline{y}+1}} \int_0^T t^{\overline{y}} e^{-K_i t} \, dt + \frac{\lambda_i^{\overline{x}+1}}{\overline{x}! K_i} \int_0^T t^{\overline{x}} e^{-K_i t} \, dt$$

(Using Claim 5)

$$= \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{K_i} T + \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{(K_i)^2} \left(e^{-K_i T} - 1\right) - \sum_{\overline{y}=1}^{\overline{x}-1} \frac{\lambda_i^{\overline{x}} \mu}{\overline{y}! K_i^{\overline{x}-\overline{y}+1}} \frac{1}{(K_i)^{\overline{y}+1}} \left(\overline{y}! - \Gamma(\overline{y}+1, K_i T)\right)$$

$$+ \frac{\lambda_i^{\overline{x}+1}}{\overline{x}! K_i} \frac{1}{(K_i)^{\overline{x}+1}} \left(\overline{x}! - \Gamma(\overline{x}+1, K_i T)\right)$$

$$= \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{K_i} T + \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{(K_i)^2} \left(e^{-K_i T} - 1\right) - \sum_{\overline{y}=1}^{\overline{x}-1} \frac{\lambda_i^{\overline{x}} \mu}{K_i^{\overline{x}+2}} + \sum_{\overline{y}=1}^{\overline{x}-1} \frac{\lambda_i^{\overline{x}} \mu}{K_i^{\overline{x}+2}} \left(\frac{\Gamma(\overline{y}+1, K_i T)}{\overline{y}!}\right)$$

$$+ \frac{\lambda_i^{\overline{x}+1}}{K_i^{\overline{x}+2}} - \frac{\lambda_i^{\overline{x}+1}}{K_i^{\overline{x}+2}} \left(\frac{\Gamma(\overline{x}+1, K_i T)}{\overline{x}!}\right)$$

$$= \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{K_i} T + \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{(K_i)^2} \left(e^{-K_i T} - 1\right) - (\overline{x}-1) \frac{\lambda_i^{\overline{x}} \mu}{K_i^{\overline{x}+2}} + \frac{\lambda_i^{\overline{x}} \mu}{K_i^{\overline{x}+2}} \sum_{\overline{y}=1}^{\overline{x}-1} \left(\frac{\Gamma(\overline{y}+1, K_i T)}{\overline{y}!}\right)$$

$$+ \frac{\lambda_i^{\overline{x}+1}}{K_i^{\overline{x}+2}} - \frac{\lambda_i^{\overline{x}+1}}{K_i^{\overline{x}+2}} \left(\frac{\Gamma(\overline{x}+1, K_i T)}{\overline{x}!}\right)$$

$$= \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{K_i} T + \left(\frac{\lambda_i}{K_i}\right)^{\overline{x}} \frac{\mu}{(K_i)^2} \left(e^{-K_i T} - \overline{x}\right) + \frac{\lambda_i^{\overline{x}} \mu}{K_i^{\overline{x}+2}} \sum_{\overline{y}=1}^{\overline{x}-1} \left(\frac{\Gamma(\overline{y}+1, K_i T)}{\overline{y}!}\right) + \frac{\lambda_i^{\overline{x}+1}}{K_i^{\overline{x}+2}} \left(1 - \frac{\Gamma(\overline{x}+1, K_i T)}{\overline{x}!}\right)$$

State $\underline{M}$

$$\int_0^T \overline{\pi}_i^t(M) \, dt = \int_0^T \left(1 - \sum_{\overline{x}=0}^{M-1} \overline{\pi}_i^t(\overline{x}) \right) dt = T - \int_0^T \sum_{\overline{x}=0}^{M-1} \overline{\pi}_i^t(\overline{x}) \, dt = T - \sum_{\overline{x}=0}^{M-1} \int_0^T \overline{\pi}_i^t(\overline{x}) \, dt$$

$\square$

Claim 5.

$$\int_0^T t^{\overline{x}} e^{-K_i t} \, dt = \frac{1}{(K_i)^{\overline{x}+1}} \left(\overline{x}! - \Gamma(\overline{x}+1, K_i T)\right)$$

Proof.

$$\int\limits_{0}^{T} t^{\overline{x}} e^{-K_i t}\, dt = \int\limits_{0}^{K_i T} \left(\frac{t'}{K_i}\right)^{\overline{x}} e^{-t'}\, \frac{dt'}{K_i} \text{ (Taking } t' = K_i t)$$

$$= \frac{1}{(K_i)^{\overline{x}+1}} \int\limits_{0}^{K_i T} (t')^{\overline{x}} e^{-t'}\, dt' = \frac{1}{(K_i)^{\overline{x}+1}} \left(\int\limits_{0}^{\infty} (t')^{\overline{x}} e^{-t'}\, dt' - \int\limits_{K_i T}^{\infty} (t')^{\overline{x}} e^{-t'}\, dt' \right)$$

$$= \frac{1}{(K_i)^{\overline{x}+1}} \left(\Gamma(\overline{x}+1) - \Gamma(\overline{x}+1, K_i T)\right) = \frac{1}{(K_i)^{\overline{x}+1}} \left(\overline{x}! - \Gamma(\overline{x}+1, K_i T)\right)$$

$\square$

D   OEF as a Stackelberg Game

Proposition 1. *Students i, j belonging to the same type $Type_l$ receive equal utility (steady state and aggregate) if they choose the same policy i.e. if $\psi^i = \psi^j (= \psi^l)$ for students $i, j \in Type_l$ then,*
$$U_\Pi^{l,i} = U_\Pi^{l,j}, \quad U_T^{l,i} = U_T^{l,j}$$

Proof. Given: $\psi^i = \psi^j \Rightarrow \psi_b^i = \psi_b^j \forall b : 1 \leq b \leq w$.

Also, $\overline{R}_{\overline{\Pi}}^{l,i} = \overline{R}_{\overline{\Pi}}^{l,j}$ if $\lambda_i = \lambda_j = \Delta_b$ (See Claim 6).

$$\therefore \overline{R}_{\overline{\Pi}}^{l,i} \phi_a \psi_b^i = \overline{R}_{\overline{\Pi}}^{l,j} \phi_a \psi_b^j$$
$$\Rightarrow D_{a,b}^{\Pi,l,i} \phi_a \psi_b^i = D_{a,b}^{\Pi,l,j} \phi_a \psi_b^j \qquad \forall a \ \forall b$$
$$\Rightarrow \sum_{a=1}^{v} \sum_{b=1}^{w} D_{a,b}^{\Pi,l,i} \phi_a \psi_b^i = \sum_{a=1}^{v} \sum_{b=1}^{w} D_{a,b}^{\Pi,l,j} \phi_a \psi_b^j$$
$$\Rightarrow U_\Pi^{l,i} = U_\Pi^{l,j}.$$

Following a similar line of proof we can also prove that $U_T^{l,i} = U_T^{l,j}$. $\square$

Claim 6. *The rewards (steady state and aggregate) received by the students i, j belonging to a $Type_l$ is same if $\lambda^i = \lambda^j$. Also the rewards (steady state and aggregate) the instructor receives w.r.t. the students i, j belonging to a $Type_l$ is same if $\lambda^i = \lambda^j$. Thus $\forall i, j \in Type_l$, if $\lambda^i = \lambda^j$ then:*
$$\overline{R}_{\overline{\Pi}}^{l,i} = \overline{R}_{\overline{\Pi}}^{l,j}, \quad \overline{R}_{\overline{\Pi}}^{I,i} = \overline{R}_{\overline{\Pi}}^{I,j}, \quad \overline{R}_T^{l,i} = \overline{R}_T^{l,j}, \quad \overline{R}_T^{I,i} = \overline{R}_T^{I,j}$$

Proof. This is trivially true as these quantities are equal if we put $\lambda_i = \lambda_j$ for students i, j belonging to $Type_l$ in the expressions given in Theorems 7 and 8. $\square$

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

Proof. Given students $i, j \in Type_l$ and considering the matrices $B^{\Pi,I,i}$ and $B^{\Pi,I,j}$:

- For entries $B_{a,b}^{\Pi,I,i}$ and $B_{a,b}^{\Pi,I,j}$, $\lambda_i = \lambda_j = \Lambda_b$ thus $\overline{R}_{\overline{\Pi}}^{I,i} = \overline{R}_{\overline{\Pi}}^{I,j}$ (Claim 6). Also $c_i = c_j \ \forall i, j \in Type_l$ (as discussed in section 5.2).

- Thus, $B_{a,b}^{\Pi,I,i} = c_i \overline{R}_{\overline{\Pi}}^{I,i} = c_j \overline{R}_{\overline{\Pi}}^{I,j} = B_{a,b}^{\Pi,I,j}$.

Thus $B^{\Pi,I,i} = B^{\Pi,I,j}(= B^{\Pi,I,l} say) \ \forall i, j \in Type_l$. Using similar arguments we can also prove that $B_{a,b}^{T,I,i} = B_{a,b}^{T,I,j}(= B^{T,I,l} say) \ \forall i, j \in Type_l$.

$\square$