

Local Preferential Attachment of Posting Behaviour on MOOC Discussion Forum

Xiaoguang Zhu^{*}, Huining Song, Yao Zhu

School of Management Engineering, Xuzhou University of Technology, Xuzhou, China

Email address:

zhu_xg@xzit.edu.cn (Xiaoguang Zhu)

^{*}Corresponding author

To cite this article:

Xiaoguang Zhu, Huining Song, Yao Zhu. Local Preferential Attachment of Posting Behaviour on MOOC Discussion Forum. *International Journal of Education, Culture and Society*. Vol. 6, No. 5, 2021, pp. 167-175. doi: 10.11648/j.ijecs.20210605.12

Received: August 25, 2021; **Accepted:** September 18, 2021; **Published:** September 29, 2021

Abstract: Online education has become a popular education paradigm and foundation for future educations. Compare to traditional offline classes, online education could support massive participants, but also confronting engagement and retention problems. Many studies have analysed participant's behaviour to enhance the engagement concerning social, cognitive and technical factors, such as 1) knowledge, cognitive process and social dimensions of micro-collaborations; 2) effects of rating visualization on reading behaviour. However, fewer studies have concerned the organization of discussion forum on posting behaviour analysis. Therefore, this study explores patterns of posting behaviours with consideration of forum organization. In detail, this study applied local preferential attachments (LPA) on the posting behaviour of MOOC discussion forum, and discovered their patterns concerning social, emotional and technical factors. The LPA model will measure message attributes on thread page as a factual local environment of posting behaviour. With statistical test on collected data, we validated the significance and linearity of the LPA on measurements of responses number and message length. Results indicated that: 1) most LPA of posting behaviour are significant; 2) comment behaviour expressed sub-linearity on responses number, and others expressed super-linearity; 3) posting behaviour was affected by compound factors of self-identity, knowledge exchange and community citizenship. Their significant preferences suggested that: 1) forum organization is significant to model posting behaviour; 2) re-organize threads and mark message status could improve engagement of participants.

Keywords: MOOC, Massive Open Online Course, Discussion Forum, Local Preferential Attachment, Posting Behaviours, Knowledge Exchange, Community Citizenship

1. Introduction

MOOC discussion forum is often an open, optional and loosely structured platform [24]. In the flexible platform, participants are motivated by psychological and cognitive factors, such as self-identity, reciprocity and preferential attachment. Those factors were studied under perspectives of knowledge sharing and social networking. For example, open reciprocity was analysed from factors of self-identity, social norm, trust and altruism [6, 8], and preferential attachment was characterized through phenomenon of super-poster and extended thread [3]. Recent studies of discussion forum were concentrated on global networking and content analysis. Networking behaviours were modelled with ERGM and SIENA methods [16, 24, 11]. Usually, following and posting actions will form a forum network. Content analysis was

applied to analyse activities of participants, such as engagement, retention and responding to a message [5, 10]. From those analyses, a posting action could be labelled as social, cognitive or affective behaviours. The general purposes of recent studies are two sides. One purpose is to decrease the burden of large volume content, hence, it's necessary to get response patterns for the classification of messages. Another purpose is to find out reasons which enhanced engagement and completion of courses, such as increase feedback to retain participants [19, 10].

Except for the discoveries of online course engagement, heterogeneity and ever-growing volume still challenging the efficiency of knowledge exchange [4, 19]. That is, new characteristics appeared as a consequence of platform evolution, such as novel organization and compound factors. For example, the EDX platform has accented forum guidance

and attached questions under course videos. Of compound factors, course platform integrated functions of open course initiative, social platform and information diffusion. Deng *et al.* also concluded that integrated functions of online education have been stabilized with dependencies on social learning, intrinsic motivation and network externality [6]. Therefore, it's necessary to discover patterns under the novel characteristics.

To discover novel patterns under the characteristics, we proposed hypotheses with local mechanism of networking analysis. Conform to platform organization, we measured the LPA with variables of message length and responses number. The reason to apply LPA is that participants are confronting local pages when posting a message. As most course syllabus designed, lectures and assignments are usually conditioned on weeks and themes [18]. On another side, platforms haven't aggregate messages in a global vision. Hence, with high probability, participants will only consider local information when posting a message, such as an answer or comment. To accommodate the changes of MOOC platform, we propose LPA model to measure the participants' propensity and growth pattern under the organization and compound factors.

In summary, discussion forum is an important component to attract participants through cognitive and social factors. This phenomenon guided previous studies to discover patterns of knowledge exchange and psychological activity as a primary theme of online education studies. In those patterns, preference attachment has been modelled in global vision, other than thread conditioned. However, posting behaviours are usually undertaken on a local page. Therefore, we propose the LPA model to discover the propensity of participants when posting messages.

Our study contributes to online course engagement with two points. Firstly, we introduced LPA modelling to posting behaviour and validated its utility. Secondly, we discovered patterns of preference under compound motivations of self-identity, knowledge exchange and community citizenship. In nutshell, we modelled the posting behaviour under the novel characteristics of MOOC platform.

2. Theory and Hypothesis

This paper constructed measurements to test significance of LPA of posting behaviour. To explain results of the tests, this section introduces theories and procedures of local preferential attachments.

Measure LPA requires an assumption that the local page affects posting behaviour independently. The rationale of this assumption is as follows. Firstly, Posting behaviour involves large cognitive investment [14, 5]. As a consequence, large volume of message will impose participants to concentrate on partial information when posting responses. Secondly, platform also divided messages into many web pages. Hence, the partial information roughly equals a single web page. For example, after a lecture, a participant may answer a question of this lecture, without considering questions of other

lectures. Except for the lecture, questions also could be organized along assignments or weeks. Those flexible organizations affected activities of participants [19, 3, 2]. For example, an online discussion thread grows either one of three patterns: short thread pattern, extended thread pattern with an elongated structure or a split thread pattern with a broad structure [3]. And, rating visualizations could affect navigation and learning of participants [2].

Given the organization and assumed context of single web page, we measure LPA with two theoretical foundations. One is local preferential attachment when modelling growth of network. Another is mechanism of knowledge exchanging which composed the local mechanisms of posting behaviour.

2.1. Local Preferential Attachment

Preferential attachment is a reasonable model to generate scale-free network, such as online social network and citation network [22, 17]. In those networks, new connections will prefer nodes with high in-degree or other attributes. In another word, when network is growing, node will get new connection with a probability that proportional to its attribute. For example, journals from the same discipline are more likely to form citation links than journals from different disciplines [16]. With integration of social and cognitive factors, activities on discussion forum may prefer special message attributes as well [23].

Compare to the preferential attachment on global network, local mechanisms were improved network modelling with likelihood estimation [20, 22]. With assumption that network should be preferentially attached to already well-connected nodes, extra local rules may stabilize the simulated degree distribution [22]. Generally, local mechanism of preferential attachment is settled with local structure, limited information and copying strategy [20, 22, 16]. Those strategies are reflected on course platform, i.e., corresponding to cognitive energy investment of posting behaviour. When generating a new connection, posters will not count the global distribution like the Price's model, but rely on local information to conduct a posting action.

2.2. Mechanisms and Variables

Except for the primary purpose to exchange knowledge, posting actions are also affected by forum organization and honour code which are designed for efficient knowledge exchange. Course platforms organized messages to facilitate search and response action of participants, such as organization along with weeks or assignments. That is, information in a single web page is sufficient to support a posting action, such as answer or comment. Given a single web page, such as a thread page, all messages on the page will form an ambient context to affect posting action through psychological and cognitive activities of participants. In another word, the posting behaviour of the participant could be fully affected by a single page.

Participants may post a message for various purposes, such as cardinal knowledge exchange, self-identity,

community citizenship and acquire feedback. In detail, self-identity and self-efficacy are fundamental conditions to enact posting actions and the reason to modify their posting behaviour, i.e., prefer to respond to one message than another [10]. Except for the self-identity, participants may be identified externally by community citizenship when perceiving external prestige and distinctiveness of online support communities [4]. Through opening process and self-organization, participants could influence online education platform with active knowledge exchanges, rather than take it as an auxiliary or impulsive learning platform. Feedback from discussion forum can promote the engagement of online course learners [7, 10]. In reverse, participants may respond to core questions for pursuing more feedback. In nutshell, when posting messages, participant will compare local ambient information to fulfil their needs of psychology and knowledge exchange.

Given the above mechanisms, we measure LPA with two variables: message length and responses number. The selection of the variables has two reasons. Firstly, both variables have sufficient visual significance to affect posting action of participant. Secondly, they could be extracted from archive data.

Message length will affect posting behaviour with the following rationale. Message length positively correlates to perplexity of the message, i.e., imply more information. Hence, with the expectation of reciprocity, a longer message will be considered more endeavoured when exchanging knowledge and attract feedback. That is, participants prefer to respond high-quality messages to improve their community [2]. In advance of self-identity, the preference conducted community citizenship through development of social identity [4]. If all participants prefer longer messages, then, it will get super-linearity with network externality. Therefore, as a major pattern, the variable of message length will achieve positive attachment level and super-linearity.

As a minor pattern, sub-linearity could happen for the following reasons. Firstly, preference of message length has sub-linearity, since the long messages are more compressible than short messages. Quantitatively, the coefficient of the linearity should in $[0, 1]$ if only consider the coding length of a message. That is, message could be compressed much shorter without losing information. Secondly, participant may respond to messages of a specific theme, since they have a specific cognitive level, habit and subjective perceptions [6]. In addition, themes of humour and technique issues could generate sub-linearity.

In those contrary effects, super-linearity should be in dominant status, because knowledge exchange is the essence of course platform, other than social activities. Hence, given the purposes and primary perceptions, we propose hypotheses of message length preference as follows.

Hypothesis 1. Participants prefer to respond to longer messages on a web page of discussion forum.

Hypothesis 2. For preference of message length, super-linear model has high likelihood than linear model.

More precise than the message length, responses number

could be functioned identically compare to in-degree of global networking model. That is, participant will prefer popular messages if discussion forum conform to traditional network growth pattern [9]. However, given the same effects about community citizenship, social presence and feedback acquisition, participants may prefer non-responded messages as well. Of linearity, insensitivity of large numbers will generate the sub-linearity effect. But on another side, complicated or top questions may attract more responses and present super-linearity as a consequence.

Hypothesis 3. Participants prefer to respond to messages which have lower responses number.

Hypothesis 4. For responses number, sub-linear model has a higher likelihood than linear model.

In nutshell, to achieve efficient knowledge sharing, posting actions will be affected by two opposite factors. One is herding behaviour which prefers popular messages. Another is self-identity and reciprocity which could equalize the responses. Our statistical test will explore the dominant factors of posting behaviour. To depict the factors, we also analysed messages qualitatively with typical preferences. Our measurements could indicate types of messages and participants, such as instructor to prefer equilibrium, and emotional contagion to transferring positive and negative emotions to others [21].

3. Methodology

To test the LPA, we reconstructed local pages of posting actions. For example, we could extract messages that a participant was confronting when he posts a comment. In detail, our study was performed as follows. We collect data at first and extract the hypothesized variables from the ambient context of each message, e.g., length of comments $\{C_{2,1}, C_{2,2}\} \in A^{C_2}$. Next, we test the significance of LPA of posting behaviour with relative attributes and student t-test. In addition, we analyse the linearity of preference through a comparison of likelihoods. At last, we select prominent examples to explain their conditions on message types.

3.1. Data Collection

Forum data were collected from two online course platforms edx.org, piazza.com. EDX supports institutions to create online courses with open and plenty of components. On this platform, participants could take free online courses from universities and institutions in the world. Piazza is a free online gathering place where students can ask, answer, and explore, under the guidance of their instructors. Both platforms provided typical components and guidance for knowledge exchange, such as forum, survey and honour codes. Hence, they could sufficiently support studies of posting behaviour, message structures and social activities of attachment and endorsement [1, 18]. Given the sufficiency, we select two courses on the platforms to test our hypotheses.

Table 1 presented the courses we were analysed, and the following will introduce their characteristics.

Table 1. Courses and Statistics.

Course	CS100.1X	ST18 6501x
Platform	Piazza	EDX
N. Question	4292	3306
N. Answer	5170	3268
N. Comment	10621	4322
N. Node	2740	609
N. Edge	8795	3748

Similar to school education, online courses also introduced components of syllabus, quiz, examination and discussion. Of discussion forum, the two courses generated fundamental thread types, such as greeting, quiz, assignment and their technical issues. Also, in process of knowledge exchange, networking and openness are truly diversified the participants and activities on those courses. That means, not only the syllabus but technical issues or social activities could dominate the forum of a specific lecture. For example, specific technique issue usually dominates a class on the discussion forum. Hence, we discover local preferential attachment to characterize the posting actions and explain their diversified pattern of knowledge exchange.

In discussion forums, typical activities are endorsements and postings of question, answer and comment. For each platform, a posted message was attached with meta-information of publish date and participant's id. We recognized the message and that meta-information with *HTML* tags and visual structures. The recognized data columns are denoted as follows.

1. Question ($q \in Q$) will generate a thread page, with answers ($A^q \in \mathcal{A}$) listed under the question. For an answer (A), its comments ($C^A \in \mathcal{C}$) will be appended to the answer. As a convention of discussion forum, answer should be solution of question, and comment will discuss the answer with supplement information, such as links and requirements.
2. Each message belongs to one type of $\{Q, \mathcal{A}, \mathcal{C}\}$, and will be attached with meta information of publish date and poster's id. In addition, question may have upvote score, and poster may have reputation and identifiable social status.
3. Statistical values. Essential attributes are message length and responses number. Optionally, the message may have been distributed on topics or tags.

3.2. Local Ambient Context

In discussion forum, most pages belong to question list or question detail. On question page, ambient context includes past questions. On question detail page, ambient context includes question content and past answers and comments of the question. Ambient information of our measurement excluded personal information and node attributes, such as gender, badge and location. Participants may answer a question as a result of searching behaviour or subscribed email, here, the ambient context will be a search result. Our measurement excluded these situations, and just consider question list in a time range or a specific class. In nutshell, we postulated that posting behaviour was determined by information on a single web page. And, ambient information

consists of messages on a single web page. The following is a schema of ambient information which formalized the variables of LPA. Figure 1 presented structures of two thread pages, where, each sub-figure denotes a question detail page and each rectangle denotes a message. On a web-page, Q is the title of current question, and Q^c is the question's content. A_i is the i^{th} answer of the Q . $C_{i,j}$ is the j^{th} comment of the A_i .

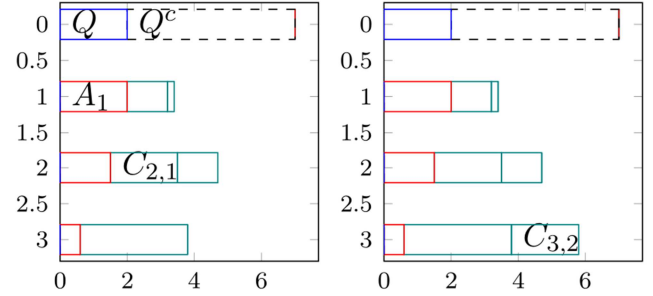


Figure 1. Local Page evolved with Visual Length.

When a new message is being generated, e.g. $C_{3,2}$ in figure 1, a local page will be expanded with a rectangle inserted into the *HTML* page. Therefore, a posting action was conditioned on two elements (M, r) , where, M denotes messages in ambient context, and r is the responded message. For each message r , we have measured the variables to test the preferential level of its posting action.

3.3. Variables to Measure LPA

Variables included message length and responses number which was measured through relative attributes of message r out of ambient messages M . Let (M, r) denote a pair (ambient messages, a responded message), and $len(\cdot)$, $deg(\cdot)$, $date(\cdot)$ denote functions of message length, responses number and publish date of message respectively. For example, $len(m)$ will return the text length of a message, and $len(M)$ will return a list of message lengths. Given a message r , we define its ambient messages (1) and relative attribute (2) as follows.

$$M = \{m \mid P(m) = P(r) \text{ \& } date(m) < date(r)\} \quad (1)$$

$$D_{M,r}^{deg} = deg(r) - mean(deg(M)) \quad (2)$$

Where $P(m)$ denotes local page of message m . The $P(m) = P(r)$ means that messages (m, r) are on the same local page, and $date(m) < date(r)$ means that message m is posted before message r , hence, it belongs to r 's ambient context. Compare to comment behaviour, answering behaviour has relatively various local environment. On platforms, *EDX* has appended questions to each lecture, hence, answering will confront a lecture page. In the contrary, for the *piazza*, we construct local question page with time windows that include 6 questions.

The relative attribute $D(M, r)$ indicated preference of participant when posting message r , and distribution of the attributes indicated LPA level of measurement. This is similar to the view of citation preference [15]. If a relative attribute is

greater than zero, we could say that posting behaviour preferred the attribute positively, i.e., preferred longer or popular messages. However, this conclusion hasn't considered various means of local pages which causes the scaling issue of the relative attribute. Hence, we also compared the distributions of measurements of responded message with random selections.

3.4. Preference and Linearity

The distribution of relative attributes may be biased by local aggregation when test LPA levels. For example, it could not discern preference $([0,2],2)$ from $([0,1,2],2)$, since they both have $D = 1$. Therefore, we also applied two-sample t-test between random selection and real distribution.

For linearity analysis, we applied Barabási-Albert model to posting behaviour with a parameter of linearity. With modified B-A models, we could approval linearity of each variable of preferential attachment. Coherent to B-A, Price's, and non-linear growth modes [12], we calculated likelihoods and added parameter α to adjust probability as follows.

$$k_i = f(i)^\alpha / \sum_{j \in M} f(j)^\alpha \quad (3)$$

$$ll = \sum_{M,r} [\log \frac{1+k_r}{\text{var}(f(M))} - \log \sum_{m \in M} \frac{1+k_m}{\text{var}(f(M))}] \quad (4)$$

where, k_m means probability with the non-linear transformation. $f(M)$ means relative attributes of the messages M , and functioned as the weight of the sum-up.

The rationality to construct equation 4 is as follows. Firstly, traditional analysis of degree distribution [13] may not precise,

because insufficient messages of the local page will cause lower statistical power, and distributions of local pages do not follow an identical event. i.e., their partition functions are different. Hence, we directly compare likelihoods of different parameters of linearity. Secondly, the smooth $1+k$ of non-linearity will impart the α an ability to indicate non-linearity. A study also assigned non-zero attachment value even to nodes of zero degree [9]. Of divide variance $\text{var}(f(M))$, we scale the calculation without alter the original likelihood with $\alpha = 1$. In another word, the division of attributes will not affect the value of the likelihood.

4. Results

This section explains results in three parts: 1) distribution of relative attributes as non-parameter analysis; 2) statistical tests to validate the significance of LPA; 3) qualitative analysis to explore message types along with different preferences. From the results, we would conclude that most hypotheses are significant in the three parts.

4.1. Distributions of the Variables

Figure 2 presented the distribution of relative attributes of $D_{M,r}$ corresponding to equations 2. Each sub-figure measured a specific attribute on a discussion forum. In detail, each row denotes a dataset, and from left to right, the columns denote number of comments, length of comment, number of answers, length of answer respectively.

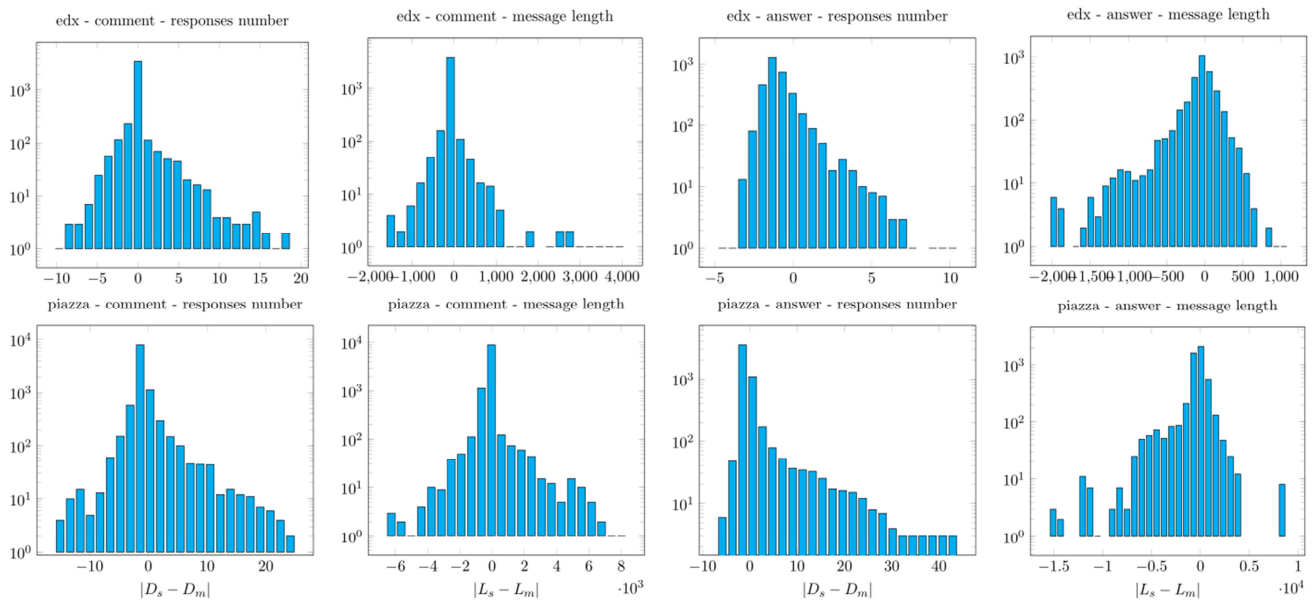


Figure 2. Deviation of Local Preferential Attachment.

The distributions are different concerning columns but have similar patterns for each measurement. Of skewness, all distributions of response number are right deviated, but message lengths are roughly left deviated. Distributions of responses number are deviated and long-tailed at right side. In the contrary, message length of answer has deviated left

side. For simplicity, table 2 aggregated the relative attributes along with their sign. In the table, each element contains two percentages: $L < 0$ and $L > 0$. If the second number is greater, then, we could speculate a positive LPA level, since message length and responses number are generally Poisson distributed.

Table 2. Global Ratio of Local Attachment.

		Msg. Length	Resp. Number
Answer	EDX	0.412 / 0.491	0.781 / 0.184
	Piazza	0.400 / 0.600	0.225 / 0.394
Comment	EDX	0.133 / 0.168	0.102 / 0.145
	Piazza	0.188 / 0.208	0.164 / 0.180

Most attributes are positively preferred as table 2 indicated, i.e., the proportion of $L > 0$ is greater than $L < 0$. The one exception, i.e., responses number of *EDX* answers, shows that community citizenship has primary effect over other factors.

Table 3. Two-sample T-test Results.

Platform	Message	Measurement	t-value	p-value
EDX	answer	Message length	3.6797	0.0002***
		Response number	-46.1760	0.0000***
	comment	Message length	1.8702	0.0614*
		Response number	1.2369	0.2161
piazza	answer	Message length	4.7021	2.62e-06***
		Response number	15.0300	3.92e-50***
	comment	Message length	3.0692	0.0021***
		Response umber	-12.6041	2.72e-36***
Note:	(*) p<0.1; (**) p<0.05; (***) p<0.01			

The significance in table 3 indicated that longer messages were positively preferred when participants posting a message. Responses number are less significant and have negative preference which derives from community citizenship. The table also corrected the result of table 2. That is, fewer deviations may have negative or in-significant LPA. Otherwise, the negative or larger preferences in table 2 are conformed to the result of table 3, such as the negative preference of response number of answer on the *EDX* platform.

The negative t-values of degree confirmed hypothesis *H3* on partial response numbers, i.e., community citizenship improved

However, the exception could be an intrinsic bias, because early published messages are encountered more which is similar to sampling bias of the Poisson process.

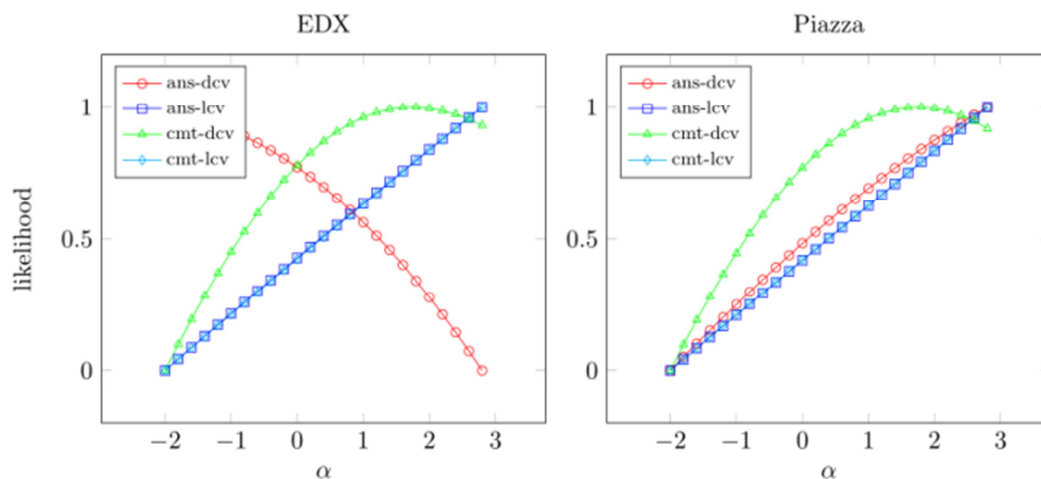
4.2. T-test and Likelihood

Table 3 is t-test results of the relative attributes. For a specific variable, the t-test was performed on two distributions of random and real responding behaviours. Similar to table 2, the t-test showed many significant LPA levels and excluded the effect of Poisson distribution on the attributes.

knowledge to exchange with feedback. However, response number of answer of piazza got a significant positive t-value. This positivity could be ascribed to search behaviour or email prescription in which the participant answered a question.

All significant positive t-values are validated hypothesis *H1*, i.e., participant will appreciate knowledge endeavour from other participants. Hence, message lengths were significantly affected posting action to respond to a message.

Figure 3 plotted curves of likelihoods concerning equation 4. In the figure, x-axis denotes parameter α , and y-axis denotes the likelihood of various models.

**Figure 3.** Likelihood along with Linearity Coefficient.

In the plotted range of α , the only local optimal showed is the degree of comment. On the contrary, message length has a large super-linearity, which means that longer messages will be appreciated by participants. However, we should notice that a message could be compressed with various levels with quite

discrepancies. For answer degree, the *EDX* expressed sub-linearity, which confirmed the non-significance of its t-test results. The *piazza* expressed super-linear. With its positive t-value, we conclude that several messages are attracted large responses with specific reasons.

4.3. Qualitative Analysis

To explore the specific reasons, this section analysed messages which were responded with a pretty high preferential level, i.e., located at terminals in figure 2. Even the specific messages are relatively less, their pattern and purpose still could reflect the social-cognitive theories which are used to construct our hypotheses. To acquire comparable message types, we sample *EDX* answers with a high preferential level of response number and plenty of messages in an ambient context. The plenty means that sum of $deg(M)$ is pretty large. Given the criteria, table 4 listed the selected

messages with four columns: degree of selected message (d_{cv}), $mean(deg(M))(d_{pm})$, question content and answer content.

In table 4, all messages are relatively urgent in different ways. In detail, messages about technical issues expressed a low preferential level, and emotion or content related messages expressed a higher preferential level. For example, approve a message implied a high level of psychological factors, such as +1, +2, and *Strangely enough*, *Thanks*, *marked as incorrect* implied emotional expression. Messages with their properties will achieve a special preferential level.

Table 4. Samples of Answers conditioned on Measurements.

d _{cv}	d _{pm}	Question	Answer
0	4.5	Hi - I signed up for the course, but was late in realizing the course already started. I have t...	Review unit 0 and the grading therein.
0	3.5	To make the answers to the multiple choice questions undebatable, it may be worth stating that t...	I think this works for any confidence interval. \nEven if it is not "tight" - because for 95% CI...
0	3	I believe there are still a couple of typos in the solution that have not yet been corrected. Co...	Fixing these. Thank you for catching these typos.
0	3	bar(X_n 2) isn't rendered correctly... Not sure what is going on...	Why do we need to refer to nn here?
0	3	I got the problem correct, but it does not seem to show up in progress. (There is no green check...	You should reference [STAFF] in this post. Click the - sign, select edit and prepend [STAFF] to...
0	3	It would be nice to have information about length of videos. I mean that now I need to run the v...	In the past Quantum Mechanics classes on edX there were the video times posted on the titles of...
0	3	Subj. Can you name good ones?:)	This book is available online for free so I like it. However, it is a little too advanced compar...
0	3	Does $XTX = I$ $T^T X = \tau I$ come from the fact that X is...	It's just an assumption that makes the homework problem easier. (As you find out by working thro...
9	2.75	Request you to kindly extend HW10 by 1 day, if possible.	+2
10	3.4	Strangely enough, but the score I got for this lecture is the lowest one in the whole course so...	... what i think is, that this course is part of Mathematics, \nin terms of what Professor repre...
8	1.12	In this question, we are given the scale parameter, not the version of the Gamma distribution wi...	+1
9	2	I have noticed that the quality of instruction in this unit, particularly in the formulation of...	I do understand the frustration, and I would be lying if I didn't feel frustrated a couple times...
9	2	Is a parameter identifiable if the model is not well specified? I think if the answer to this qu...	I got this one wrong too, but after watching the "Identifiable" video again I agree with the exp...
10	3	Request you to kindly extend HW10 by 1 day, if possible.	Since we are nearing the end of the term and staff ourselves have needed extensions, we will ext...
10	1.89	Please take a look. Have an answer in slightly different notation, but it is marked as incorrect.	used $q(\alpha/2)$
11	2	Please take a look. Have an answer in slightly different notation, but it is marked as incorrect.	The delta method does not apply here, when ' $p = 1/2$ ', we have a 0 in a denominator. Can help me...
12	2.11	Please take a look. Have an answer in slightly different notation, but it is marked as incorrect.	I used $q(0.025)$ too! Thanks
13	2.22	Please take a look. Have an answer in slightly different notation, but it is marked as incorrect.	same

Theoretically, the samples conformed to studies of behaviour patterns of self-identity, citizenship, emotional contagion in section 2. For a message, its first response may prefer self-identity or cognitive achievement, and the following may express emotional traits. That is, the response at the beginning could be reinforced with instructions [8], and the latter may prefer approval or emotional expression, such as confirmation or summarizing process [3, 1]. In the middle range, knowledge exchange related interactions are essential reasons to sustain engagement [11]. In another word, interactive participants will continuously engage the course, and then post more

messages with self-nourished interactivity.

5. Discussion

In this paper, we proposed a local attachment model to discover the novel pattern of posting behaviour in online course forum. Through t-test and likelihood ratio, we validated hypotheses about the pattern on two collected datasets. Experiment results confirmed most of the hypotheses and specificity of messages. Given the hypotheses and results, we discuss concepts of the local and preferential as follows to clarify our model.

5.1. Local Page and Comparative Analysis

The local page is an ambient context that independently affects posting behaviour of a participant. With data volume increasing, it is necessary to concentrate on local page when posting a message. On another side, the unified platform organization facilitated the networking model with the local mechanism. That is, we could find out which ambient context has affected posting behaviour of participants. Given the necessity and sufficiency, we could model posting behaviour with local preferential attachment.

When posting a message, participants will compare the attributes of messages on the local page. The final posting action will be affected by the comparison and traits of the poster. In another word, related attributes of the final posting will reflect the social-cognitive pattern of the posting behaviour. In addition, the comparison may not happen in some cases, since participants may post a message instantly or through search behaviour. This posting strategy will generate specific messages and preferential levels in our qualitative analysis.

5.2. Mechanism of the LPA

In networking behaviours, new connections are primarily affected by preference or purpose, e.g., the purpose of self-present may enact a connection to core nodes. When posting a message, the purpose of participant also determined their posting behaviours. Since purpose of participant could not be observed directly, we modelled the LPA to indicate the purpose. In reverse, our results validated that posting behaviours were affected by compound factors, such as self-identity and community citizenship.

The factors assimilated participants toward conformity and criteria of course platform. To achieve efficient knowledge exchange, participants will follow guidance, honour code and citizenship in platforms. Those factors are shaping homogeneity of participants as foundation of our hypotheses. In platforms, category and visualization of messages also affected activities of participants [1, 2]. This is a reverse way that assimilate the participants. With a pretty large investment of cognitive energy, participants will care about those endeavours. Hence, in process of responding, they will compare and enact reasonably. In detail, learner may passively mimic their peers (irrational herding) or engages in active observational learning concerning their needs (rational herding) [23].

5.3. Research Comparison

This study shared several common points with previous studies. Firstly, there is a common purpose to measure interactivity behaviour and fulfil the requirement of online education analysis [11]. Secondly, methods of statistical test, networking model and content analysis are applied to online education analysis, such as ANOVA [2], likelihood ratio [14], SIENA [24]. Thirdly, we introduced similar objects and factors to model online education, such as Piazza [1], AOD forum [3] and Facebook Group [6]. Most of those analyses are performed on units of message, thread and participant. We have resorted to classical and specified factors which are

significant on education platforms, such as community citizenship [4], cognitive engagement [5] and psychological needs [7].

This study also extended traditional methods to discover the pattern of posting behaviour concerning forum organization. Firstly, previous studies focused on single aspects of message types, response times and growth patterns on posting behaviour [1, 3]. Our study discovered compound factors that affect posting behaviour. We analysed those preferential attachments which are conditioned on message attributes. Secondly, we applied LPA [20] concerning forum organization and factual environment of posting behaviour which is different to the reading behaviour or community size with network modelling [2, 24]. In addition, we combined qualitative and quantitative analyses to validate the patterns of posting behaviours.

6. Conclusion

Through statistical tests on two datasets, we conclude our study as follows. Firstly, local mechanisms could be applied to networking analysis on the *MOOC* discussion forum. That is, we discovered significant local attachments through the LPA models. Secondly, posting behaviours are affected by compound factors which include self-identity, knowledge exchange and community citizenship. Thirdly, participants preferred messages with specific types, such as technical issues or emotional expressions.

This study validated that forum organization is usually to model participant behaviour because the organization decides the factual environment of posting behaviour. Therefore, we could improve forum organization and online learning environment with the following strategies. Firstly, we should add guidance on the format of questions and answers to make them articulated. Secondly, lecturer should harness the emotional and technical issues with quick responses. If not, those threads could increase quickly and attract attention of participants. Thirdly, we can mark the status and importance of posts to evoke community citizenship and posting behaviour of participants.

Acknowledgements

This work has been supported by the Jiangsu Modern Education Technology Project 2021-R-87014.

References

- [1] Mohan John Blooma, Jayan Chirayath Kurian, Alton Yeow Kuan Chua, Dion Hoe Lian Goh, and Nguyen Huong Lien. Social question answering: Analyzing knowledge, cognitive processes and social dimensions of micro-collaborations. *Comput. Educ.*, 69: 109–120, 2013.
- [2] Jürgen Buder, Christina Schwind, Anja Rudat, and Daniel Bodemer. Selective reading of large online forum discussions: The impact of rating visualizations on navigation and learning. *Comput. Human Behav.*, 44: 191–201, 2015.

- [3] J. C. C. Chan, K. F. Hew, and W. S. Cheung. Asynchronous online discussion thread development: Examining growth patterns and peer-facilitation techniques. *J. Comput. Assist. Learn.*, 25 (5): 438–452, 2009.
- [4] Chao-Min Chiu, Hsin-Yi Huang, Hsiang-Lan Cheng, and Pei-Chen Sun. Understanding online community citizenship behaviors through social support and social identity. *Int. J. Inf. Manage.*, 35 (4): 504–519, 2015.
- [5] Michelene T. H. Chi and Ruth Wylie. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educ. Psychol.*, 49 (4): 219–243, 2014.
- [6] Liping Deng and Nicole Judith Tavares. From Moodle to Facebook: Exploring students' motivation and experiences in online communities. *Comput. Educ.*, 68 (October): 167–176, 2013.
- [7] Jiaming Fang, Lufen Tang, Jingjing Yang, and Min Peng. Social interaction in MOOCs: The mediating effects of immersive experience and psychological needs satisfaction. *Telemat. Informatics*, 39: 75–91, 2019.
- [8] Khe Foon Hew and Wing Sum Cheung. Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educ. Res. Rev.*, 12: 45–58, 2014.
- [9] H. Jeong, Z. Nédá, and A. L. Barabási. Measuring preferential attachment in evolving networks. *Europhys. Lett.*, 61 (4): 567–572, 2003.
- [10] [Yeonji Jung and Jeongmin Lee. Learning Engagement and Persistence in Massive Open Online Courses (MOOCs). *Comput. Educ.*, 122: 9–22, 2018.
- [11] Carmel Kent, Esther Laslo, and Sheizaf Rafaeli. Interactivity in online discussions and learning outcomes. *Comput. Educ.*, 97: 116–128, 2016.
- [12] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Phys. Rev. Lett.*, 85 (21): 4629–4632, 2000.
- [13] Jérôme Kunegis, Marcel Blattner, and Christine Moser. Preferential Attachment in Online Networks: Measurement and Explanations. In *WebSci*, pages 1–10, Paris, France, mar 2013. ACM.
- [14] Saskia Metzler, Stephan Günnemann, and Pauli Miettinen. Stability and dynamics of communities on online question-answer sites. *Soc. Networks*, 58: 50–58, 2019.
- [15] M E J Newman. Prediction of highly cited papers. *Europhys. Lett.*, 105 (2): 28002, 2014.
- [16] Tai-Quan Peng. Assortative mixing, preferential attachment, and triadic closure: A longitudinal study of tie-generative mechanisms in journal citation networks. *J. Informetr.*, 9 (2): 250–262, 2015.
- [17] Camille Roth and Jean Philippe Cointet. Social and semantic coevolution in knowledge networks. *Soc. Networks*, 32 (1): 16–29, 2010.
- [18] Daniel T. Seaton, Yoav Bergner, Isaac Chuang, Piotr Mitros, and David E. Pritchard. Who does what in a massive open online course? *Commun. ACM*, 57 (4): 58–65, 2014.
- [19] Ayse Saliha Sunar, Su White, Nor Aniza Abdullah, and Hugh C. Davis. How learners' interactions sustain engagement: A MOOC case study. *IEEE Trans. Learn. Technol.*, 10 (4): 475–487, 2017.
- [20] Alexei Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E*, 67 (5): 1–15, 2003.
- [21] Inder M Verma. Editorial expression of concern: Experimental evidence of massive-scale emotional contagion through social networks. *Proc. Natl. Acad. Sci. U. S. A.*, 111 (29): 10779, 2014.
- [22] Li-Na Wang, Jin-Li Guo, Han-Xin Yang, and Tao Zhou. Local preferential attachment model for hierarchical networks. *Phys. A Stat. Mech. its Appl.*, 388 (8): 1713–1720, 2009.
- [23] Wei Wang, Lihuan Guo, and Rui Sun. Rational herd behavior in online learning: Insights from MOOC. *Comput. Human Behav.*, 92: 660–669, 2019.
- [24] Jingjing Zhang, Maxim Skryabin, and Xiongwei Song. Understanding the dynamics of MOOC discussion forums with simulation investigation for empirical network analysis (SIENA). *Distance Educ.*, 37 (3): 270–286, 2016.