How Effective is Your Facilitation? Group-Level Analytics of MOOC Forums

Oleksandra Poquet
University of South Australia
Delft University of Technology
Adelaide, Australia
sspoquet@gmail.com

Shane Dawson
Teaching Innovation Unit
University of South Australia
Adelaide, Australia
shane.dawson@unisa.edu.au

Nia Dowell
Institute for Intelligent Systems
University of Memphis
Memphis, USA
ndowell@memphis.edu

ABSTRACT
The facilitation of interpersonal relationships within a respectful learning climate is an important aspect of teaching practice. However, in large-scale online contexts, such as MOOCs, the number of learners and highly asynchronous nature militates against the development of a sense of belonging and dyadic trust. Given these challenges, instead of conventional instruments that reflect learners' affective perceptions, we suggest a set of indicators that can be used to evaluate social activity in relation to the participation structure. These group-level indicators can then help teachers to gain insights into the evolution of social activity shaped by their facilitation choices. For this study, group-level indicators were derived from measuring information exchange activity between the returning MOOC posters. By conceptualizing this group as an identity-based community, we can apply exponential random graph modelling to explain the network's structure through the configurations of direct reciprocity, generalized exchange, and the effect of participants demonstrating super-posting behavior. The findings provide novel insights into network amplification, and highlight the differences between the courses employing different facilitation strategies. Direct reciprocation was characteristic of non-facilitated groups. Generalized exchange was more prominent in highly facilitated online communities with instructor's involvement. Super-posting activity was less pronounced in networks with higher generalized exchange, and more pronounced in networks with higher direct reciprocity.

CCS Concepts
Applied computing →Education→E-learning

Keywords
MOOCs, forum, facilitation, indicators of social activity, ERGM.

1. INTRODUCTION
With the massification and commercialization of higher education, universities place more demands on educators to deliver quality teaching to more students at lower costs. As a result, educators need to adapt and modify their pedagogical approaches to better accommodate a significantly larger and more diverse group of students. Well-designed assessment to some extent can help teachers determine if their efforts to help students learn were effective. However, student performance data alone, does not provide sufficient information about other aspects of the learning and teaching context. An alternate approach is required to explore if a teacher's implemented facilitation processes were successful in promoting interpersonal relationships and peer-to-peer learning.

Various instruments have been designed to capture the evolution of in-class social relations related to the positive impact on learning. Examples include the community of inquiry questionnaire [45], surveys of social presence [34] and sense of belonging [46]. These instruments evaluate the state of trust and community often at the group-level as the ultimate outcome of interpersonal relationship formation.

The importance of trust-based relationships has been associated with small online and face-to-face classes where the group boundaries are fixed. However, there remains a significant challenge in leveraging the benefits of such interpersonal relationships in large learner cohorts particularly in the non-formal contexts. In massive open online courses (MOOCs) learner participation in social activities is intermittent [55], and early interactions can amass to a level of chaos [3] that impedes an individual's propensity for developing relationships. For instance, according to Gillani, MOOC forums 'assemble and disperse as crowds' [10]. Additionally, as stated by Poquet et al. [39] the relatively short time frames associated with MOOC offerings also further diminish the opportunities for learners to develop interpersonal trust. Simply put, large course size, short duration of teaching and the non-formal nature of MOOCs militate against the development of community. In such instances the use of established instruments evaluating if learner-learner bonds have been forged becomes irrelevant.

The amplification of communication between participants could serve as an effective and proactive indicator for group processes. The promotion of connections between learners, concepts and artefacts is favoured by networked learning, connectivism and socio-material approaches [1, 28, 48]. For example, a connectivist approach to teaching in networked systems includes controlling the network of learners through a facilitating role – by amplifying, curating, way-finding and socially-driven sense-making, aggregating, filtering, modelling and being persistently present [49]. In this context, the role of the teacher becomes strongly aligned with promotion and facilitation of network's interconnectedness. Thus, an amplified network of learner communications may indicate if a structure and a climate conducive to egalitarian participation has evolved.

This present study makes use of network indicators to explore the state of social activity in an open online course. Through these

Conference'10, Month 1–2, 2010, City, State, Country.

DOI: http://dx.doi.org/10.1145/12345.67890

indicators the study examined amplification of communication within the network of learners in ten discrete MOOCs. Exponential random graph modelling (ERGMs) was applied to understand the extent of direct and generalized reciprocity and the effects of participants posting activity within each network. ERGMs required a theoretically-driven hypothesis to be interpreted meaningfully. Thus, based on our prior work we conceptualized groups of regularly posting forum users, i.e. forum cores, as identity-based online communities. Communication within the communities were abstracted into network graphs for ERGM analysis. Results highlighted the differences in propensity for tie formation and the probability of information exchange across networks of different size and course design. The examined MOOCs were conducted in STEM disciplines, and delivered by the Delft University of Technology via the edX platform in 2013-2015.

2. REVIEW OF LITERATURE

2.1 Analytics for Group-Level Social Activity

The social fabric of a learning group evolves from the structures and processes created as the collective completes the assigned work. The development of effective group strategies requires teachers to adopt strong facilitation skills and techniques. In this context, facilitation skills refer to the ability to design and manage learning group structures and processes, and employ appropriate and timely strategies to minimize the common problems that arise when people work together [29]. Facilitation skills, learning design and support of knowledge construction are all required to establish an effective online teaching program [7]. However, facilitation can be seen as distinct from the other two. Learning design is geared towards teachers decisions about learning tasks, course structure and assessment [6]. The support of knowledge construction, or 'content facilitation' [13], refers to a teacher's ability to promote students' epistemic fluency. Facilitation is content-neutral in that it promotes a structure and a culture of participation built on individual dispositions to the group and one another. Measuring learning outcomes that result from effective facilitation is particularly challenging because they fall within the socio-emotional domain.

Group-focused analyses can provide novel insights into how social activity unfolds. For example, Huang et al [24] examined the health of forums at the group level in 44 Coursera MOOCs in relation to super-posting behaviours of its learners. The authors found that overall higher activity from super-posters is positively correlated with higher activity from other forum users, concluding that the super-posters do not 'drown out the silent majority' (p.118). It should be noted, however, that such an evaluation of the health of the forum was limited to the correlation of individual counts of content-related messages between different types of learners. The study is a reflective of how social activity on forums is approached in MOOC research to date.

In current research of platform-based MOOCs, most indicators of social activity are not well suited to evaluate the state of social activity in a group. Analyses have been predominantly conducted at the level of the individual, often as a complement to academic performance (grades). Numerous studies have analysed the association of a number of forum posts in relation to an individual's performance [12]. For example, Kizilcec et al. [31] demonstrated that completing learners who attempted a majority of the assessments were also more active on the forum. Vu et al. [52], however, found that the relationship between learning performance and social interactions is not bi-directional. Although high learning performance was positively correlated with active social interactions, active social interactions were not necessarily associated with performance. The importance of these analyses lies in grasping the value of social activity for individual student certification, however, they offer little insight into the group-level forum dynamics.

Studies measuring individual engagement with others in the form of social capital offer an alternative to evaluating the role of social activity through post counts. Such an approach is based on quantifying an individual learner's position within the network of posters [8, 25, 26]. However, such an approach does little to reveal how the network emerges from the collective contributions. To explain, although individual learning outcomes such as social capital do provide information about social activity, they are not sound indicators of group-level activities.

Where social activity in MOOCs has been evaluated at the group-level, the evaluations have tended to privilege the content-dimension of interaction. For instance, Hecking et al. [21] encapsulated social interaction within content-related posts of the network into blockmodels describing the overall network pattern. Or, Kellogg et al. [30] reported knowledge construction within the network of learners through qualitative analysis of learner posts. Similarly, Wang et al. [53] identified and evaluated groups in the course based on higher order thinking behaviours reflected through posted text. These examples demonstrate that the socioemotional dimension of learning is often taken for granted. As argued by Kreijns et al. [33], when a need for the socioemotional is recognized, the dimension is examined in the context of a content-loaded learning task or as targeting cognitive processes.

2.2 Social Network Analysis in MOOCs

Social network analysis (SNA) is widely used for evaluating group-level social activity. SNA as a methodology combines learner agency with the context where it operates. Thus, it offers group-level insights into social activity in MOOCs. Early on Ferguson & Shum [9] suggested the use of SNA for investigating interpersonal relations mediated by social platforms. They proposed to examine social relations through the prism of strong and weak ties – theoretically loaded terms borrowed from social science theory. This proposition aligns with the suggested focus on inquiries into learning ties by Haythornthwaite & de Laat [20]. However, any inferences about group-level activity in MOOCs drawn from current SNA research in learning analytics needs to be enacted with caution.

On the one hand, SNA requires its methodological decisions to be theoretically driven. Network research is relational, that is, its main unit of analysis is a theoretically defined relation between two actors. The meaning assigned to this unit of analysis directly impacts the interpretations of the results. In short, a network is a representation of a phenomenon, and consequently, this phenomenon needs to be conceptually represented in the data [2]. In practical terms this is done by providing justification for inclusion and exclusion of ties and edges, as well as drawing network boundaries. According to Laumann, Marsden and Prensky [35], errors in defining these boundaries reach beyond the consequences of slightly biased estimates of population means, proportions, or inefficiency in statistical estimation. Flawed boundary specification may lead to "the fundamental misrepresentation of the process under study, since …errors of omitting one actor may distort the overall configuration of actors in the system and render the entire analysis meaningless" (p.19).

On the other hand, there is a lack of theorising social processes in open online courses. As a result, there is a divergence in the research resulting from the differing conceptualizations of network ties. This largely stems from how post-reply structures are interpreted. For instance, researchers examining Twitter networks in MOOCs analyzed communication represented through person-to-person information flow [15, 50]. Twitter communication interactions are possibly the least ambiguous to interpret as they are by design identical with the name networks where there is little doubt as to 'who spoke to whom' [16]. Post-reply communication structures appear to be replicated in studies of platform-based MOOCs delivered via edX and Coursera. However, the approaches adopted to analyse the forum data differ. For instance, Joksimovic, et al., [27] used a conventional directed post-reply structure commonly used in online education forums: "if author A2 replied to a message posted by author A1, we would add a directed edge A2->A1. Further, if A3 posted a comment on A2's post, we would include A3->A2 edge as well". In contrast, Brown and colleagues [4], [5] used a post-reply network structure based on communication going from one person to many prior posters. Gillani, et al., [11] adopted another alternate approach in a Coursera MOOC. In this instance, the authors represented communication through an edge connecting "two learners simply if they co-posted in at least one discussion thread". The diversity in how researchers are conceptualising information/ communication flow as network structures impedes our capacity to make broader inferences about group-level learning activity. Hence, to draw meaningful interpretations from graph analytical techniques employed in SNA of MOOCs, it is imperative to theorise and conceptualise the phenomenon that is being studied.

3. THE CURRENT STUDY

3.1. Our Approach

The aim of the current study is to develop indicators to evaluate social activity in MOOCs, particularly in relation to forum facilitation. Given that interpersonal relationships at scale may be unattainable, we propose to examine the amplification of communication within the networks of MOOC learners.

In our earlier research we argued that teaching online approaches are derived from formal educational contexts, and there is a strong need for new analytical frameworks that are distinct for MOOC settings [39]. While open and formal online learning appear similar, processes in formal online modes are streamlined by the strict starting dates and learners' motivation to complete. The formal boundaries and learner motivations lie in stark contrast to the more informal and open MOOC contexts. Yet, both research and practice use theoretical and methodological frameworks inherited from formal settings to analyse open online courses. To overcome the challenge of compatibility of the tools with the object of analysis, social processes emerging on MOOC forums first need to be examined for forum sub-groups comparable in dynamics to formal education settings. In contrast to many MOOC learners who drop into the forums and leave, persistent forum contributors experience a greater sense of continuity and established history in their relationships typical for social groups [40]. Thus, a sub-set of forum participants, termed as the *forum core*, was established.

The *forum core* are posters who contributed to the forum in at least any three weeks of the course and received replies to their posts within the same course week. The underlying characteristic of this sub-group is repeated presence and timeliness of the reciprocated communication exchanges. The number of posts does not distinguish this group from other posters. Some forum core members could have as few as four to five posts, and some non-forum core posters could have over a hundred.

To meaningfully capture indicators of participation patterns we conceptualized *forum core*, i.e. a group of returning MOOC posters as identity-based community. Then, building on prior research examining the properties of network formation in social exchange networks, we inquired about the effects different facilitation strategies may have on network configurations. The underlying theoretical assumptions were evaluated by statistical network analyses applied to ten cases of MOOCs. The following section offers an overview of the prior work and relevant theoretical and empirical literature that shaped the research question.

3.2 Identity-Based Communities

Education research has tended to define community development within two broad categories – community formed through a common-identity or formed via a common-bond. The differences between these categories lie within the origin of the group-attachment. One-to-one attachment among members is a consequence of people *liking* each other. When approaching attachment from a bond-based perspective, the focus is therefore on interpersonal, dyadic relationships of trust. In contrast, common-identity attachment is a result of *self-identification with group's purpos*e as in topic-based groups, such as sports team or a school newspaper. These differences between common-identity and common-bond communities are best captured by what happens when a member leaves. When attachment is motivated by identity, other members are perceived as interchangeable, and turnover of membership does not impact individual behaviour as much as in common-bond communities where members will leave following their friends.

A common-identity approach has previously been applied to informal online groups [41, 47, 51]. Postmes, Spears & Lea [44] argued that the power of anonymity in groups where individuals do not have cognitive representations of other individuals could enhance the salience of a group's identity. The authors suggest that 'paradigms in which group members do not meet face-to-face provide precisely those conditions predicted to maximize social influence exerted by social norms and social identities' (p. 7). Furthermore, in a series of experiments, Ren et al. [42] demonstrated that facilitation of identity-based attachment had stronger effects on the frequency of engagement than conditions targeting interpersonal attachment development. Therefore, defining a community through identity-based attachment of one-to-many is of high relevance in educational settings such as MOOCs where the scale and non-privacy of the environment aggravate the development of bond-based relationships.

In light of identity-based community theory, this present study approached *forum core* as a social entity where attachment is dictated by self-identification with a group without necessarily investing in specific person-to-person relationships. Examination of the discourse produced by *forum core* indicated traces of social processes that *forum core* posters engaged in [40]. More specifically, through qualitative analysis of forum content, Poquet [38] identified the presence of discourse negotiating situational norms, shared domain of practice, i.e. MOOC forum, and shared experience, and, MOOC learning and assessment. Furthermore, in a complementary study of the perceptions among those contributing regularly Poquet et al. [39] found that forum core participants reported the lack of dyadic trust and dyadic familiarity in interpersonal perceptions but presence of higher group cohesion perceptions. In short, *forum cores* in some courses engaged in collective social processes, and developed perceptions of group cohesion without underpinning interpersonal trust. This empirical evidence allows to conceptualize the forum core as an identity-based community based on information exchange.

3.4 Network Exchange Patterns in Online Communities

Insights into the baseline properties of similar networks are required to transition from theoretical analysis of forum core as identity-based communities to applied analysis of the amplification of its communication exchange. Studies of electronic networks of practice offer information about network topography in identity-based communities that help modelling forum core participation structure.

Electronic networks of practice are conceptually aligned with the major premises of identity-based communities. They are informal groups that share knowledge about specific subject of interest from photography groups on flickr to Wikipedia contributors to software developers blogging community. According to Wasko et al., [54], participants of public electronic networks reported trust in the quality of information from active members (i.e. attachment to the group), and that those actively participating were perceived as trustworthy (i.e. belief in the group), thereby eliciting a stronger motivation to continue participation (i.e. commitment to participating in group's activity). Faraj & Johnson theorised such online communities as social exchanges between participants situated in a network context. That is, regardless of the content of exchange, the interactions are more than just information queries, but also have social nature and social purpose. In other words, in their activities posters are conscious of the potential use by readers, as well as current and future contributors through what becomes a part of shared history. Simply put, electronic networks of practice appear to represent a particular kind of an identity-based community.

The work of Faraj & Johnson served to empirically validate the consistent and predictable network exchange patterns that emerge from divergent motivations. They proposed three network patterns that characterize the formation of the network of practice on the micro-level: direct reciprocity (A replies to B because B helped A in the past), indirect reciprocity (B helped to A, and thus A will help C), and preferential attachment where new actors choose to interact with already well-connected actors. Through a large-scale longitudinal research of Yahoo! Bulletin boards, they found a significant positive high propensity of direct reciprocation (A to B to A), significant positive low propensity for indirect exchange happening (A to B to C), and a negative propensity for preferential attachment. The negative propensity for preferential attachment has also been supported in studies of MOOC forum networks [27, 30]. In summary, network configurations of reciprocity lend themselves nicely to modelling baseline network properties in forum core networks for further examination of the differences between them in different facilitation contexts.

3.4. Focus of the Study

Building on the discussed theoretical and empirical literature, MOOC forum cores are assumed to reflect natural patterns occurring in identity-based communities characterized by network exchange. Accordingly, forum cores that emerge and evolve in courses with pedagogical interventions constituted by moderation strategies would demonstrate different dynamics, varying along with the intensity of facilitation. These differences in network structure in courses with moderate and higher moderation would then indicate the effects of intended forum strategies. In line with this argument, the following research question was posed:

Are there differences in patterns of direct and generalized reciprocity in forum core networks under different facilitation strategies?

To address this research question forum core communication networks were analysed using models of social selection based on the *p**class models [44].

4. METHODS

4.1. Network Construction

Forum core networks were constructed based on the premise that they represented identity-based communities. That is, the analyses were constructed from the forum core posters that were committed to participating in forum's activity, and driven by the mix of altruistic and selfish reasons to engage in information exchange. Their commitment was captured by the extended participation criterion that distinguished forum core posters from intermittent posters.

Network ties were non-valued and directed representing the direction of the service offered. The ties were directed to all the posters in the same thread who preceded the actor based on timestamps. If A posted, B replied, and C commented, and D replied, each of the subsequent actors would have a directed tie to everybody else prior to them. That is, B -> A, C-> B, C->A, D->C, D->B, D->A. Such tie inclusion was based on the principle of collective reciprocity where an individual contributed information to the group, rather than provided a service to an identified individual.

The network boundaries were set around the course delivery time: between the week the first video lecture was released, and the week when the last lecture in the course was released. As communication between actors can potentially continue well past the course lectures due to exams or assignments, any established closing date for the forum analysis would be arbitrary. The discrete temporal limits applied in this study made the comparison of courses more feasible.

4.2 Data Description

This present study analysed the forum interactions evolving from ten (10) STEM courses delivered by the Delft University of technology via edX platform in 2013-2015. Due to the evaluative nature of the study, the courses were de-identified. Table 1 offers an overview of the disciplines, sizes of posting cohort against the forum core, course duration and qualitative description of the forum facilitation strategies.

The dataset included five large MOOC with forum core higher than one hundred people, and five smaller courses with forum cores ranging from 23–70 people. Three courses were highly facilitated, i.e. instructor was among active posters, along with designated staff members and TAs. Five courses had moderate facilitation, i.e. although staff and TAs helped the information flow, instructor was not involved with the forum. Finally, two courses had low facilitation, i.e. nobody moderated the forums except one or two staff announcement were posted.

Table 1. Course overview with forum facilitation strategies.

Course	Area	Posters	Forum Core	Course Duration	Forum Facilitation
A	Soft Systems	752	48	10	Moderate
B	Engineering	4384	340	10	Low
C	Engineering	668	38	6	Moderate
D	Soft Systems	1859	70	6	Moderate
E	Engineering	2480	231	9	High
F	Engineering	6775	254	10	Moderate
G	Engineering	360	23	5	Low
H	Soft Systems	1255	43	5	Moderate
I	Data Analysis	2825	220	8	High
J	Computer Science	1433	161	8	High

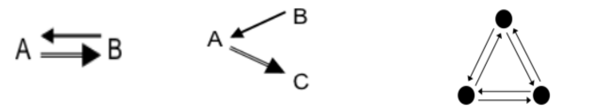
4.3 Exponential Random Graph Modelling

Exponential random graph modelling (ERGM) was used to analyse the network properties describing the structure of forum core networks. ERGM is a methodology for the analysis of social selection models. These models assume that characteristics of actors influence them to select or be selected by others as social partners [44]. These micro-level interactions resulting from social selection aggregate into group-level patterns that describe the network. Additionally, p*/ERGM presumes that multiple processes can take place simultaneously, and social networks are both structured and stochastic [36 p.10].

ERGM is a probability model for network analysis [17, 19, 36, 37]. In modelling ERGM estimates the likelihood of a parameter, i.e. theoretically formulated structural network characteristic, to occur beyond chance. This is implemented by examining the likelihood of a studied parameter to occur in a generated distribution of random networks of the same size as the network of interest. Additionally, multi-level analysis within ERGM controls for the tendency of studied parameters against one another, due to their theoretical dependency. The output includes a parameter estimate where zero indicates that the modelled effect is not different from random. Estimated parameters are considered of value after the goodness of fit is conducted upon the best fitting model.

4.4 Modelled Configurations and Effects

This study investigated three parameters in forum core networks: direct reciprocity characterized by *mutuality statistic*, generalized exchange characterized by *simmelianties statistic*, and the *effect of learner super-posting activity* on the propensity of sending and receiving ties. Figure 1 depicts three network configurations. The first two are direct and indirect reciprocity that, according to Faraj & Johnson are baseline properties of network formation in identity-based communities. We further hypothesised that forum core members who overtime co-occur within direct (A to B to A, Figure 1a) and indirect reciprocity (A to B to C, Figure 1b) will overtime form triadic network configurations representing mutual and generalized exchange that eventually could extend to a cycle if interaction should continue. Joksimovic, et al., [27] previously fitted *simmelianties statistic* in ERGM for MOOC forums, and interpreted it as Simmelian ties [32]. From the analytical perspective, we take inspiration and build on their approach. Yet, we interpret this network configuration as amplification of parts of the network due to generalized information exchange between the actors. Only direct reciprocation and generalized exchange configurations were therefore modelled.



a) Direct Reciprocity **b) Indirect Reciprocity** **c) Generalized exchange**

Figure 1. Network Features Modelled in ERGMs. Adapted from Faraj & Johnson (2010).

Markov network configuration were not used in the study to model triadic generalized exchange. Instead we use *simmelainties* statistic available in R statnet package [18]. Morris, Hunter & Handcock [37] explain that simmelian triads (Figure 1c) can overlap in terms of nodes and ties, thus *simmelianties* is rather a 'measure of the clustering of Simmelians (given the number of Simmelians).' Although social circuit parameters (gwesp) were fitted on some of the networks, they were not characteristic of all networks, and thus not used to maintain comparability in the modelled outputs. Table 2 presents the counts of modelled network configurations for each of the networks.

Table 2. Counts of network configuration in analyzed networks.

Course/ Facilitation	Nodes	Edges (*density**)	Reciprocity (*mutual*)	Ties embedded in generalized exchange (*simmelianties*)
A/Moderate	48	968	306	602
B/Low	340	4574	725	1164
C/Moderate	38	367	111	196
D/Moderate	70	713	164	264
E/High	231	3743	1003	1938
F/ Moderate	254	3059	629	1060
G/ Low	23	86	19	0
H/ Moderate	43	249	50	34
I/ High	220	2260	516	882
J/ High	161	2978	880	1734

Note: *equivalent statnet name for the parameters

Additionally, in ERGMS we controlled for the activity level of each learner within the network using the *nodefactor* parameter. To establish the overall level of activity, we applied in-reach and out-reach measures of each individual activity as suggested by Hecking, Hoppe & Harrer [22]. The authors developed measures of entropy to calculate the diversity of outgoing and ingoing relations for a node, and account for the weight of an edge, i.e. frequency of dyadic ties between two given actors. We have replicated their measure[1], and applied it to the entire forum network. We then applied k-means clustering to divide all forum posters into three groups: with highest, moderate and low forum posting activity in the entire network. These attributes of posting activity were used to control for tie propensity formation in forum core ERGMs. Table 3 presents the number of actors within each cluster. Table 3 demonstrates the number of people with different posting activity in each of the networks. Cluster 1 refers to the learners with low posting activity. Cluster 2 refers to the posters with moderate posting activity, and Cluster 3 represents hyperposters. These interpretations do not fit to describe Course C, where Cluster 3 and Cluster 4, in fact have similar activity levels, but Cluster 2 has higher in-reach measures, and Cluster 3 has higher out-reach measure.

Table 3. Number of people in the cluster representing different posting activity after k-means clustering of in-reach

[1] https://github.com/hecking/socio_semantic_blockmodelling/blob/master/scripts/centrality_over_time.R

and out-reach measures of posting activity in the entire network

Course	Cluster1	Cluster2	Cluster3
A	503	65	3
B	2998	273	6
C	381	75	50
D	856	398	1
E	1297	459	4
F	4082	40	7
G	262	58	7
H	849	61	11
I	1994	38	3
J	1317	41	4

It should also be noted that all labels are relative to one another within the course. That is, a hyperposter in one course may have similar activity numerically as the moderate poster in another course. Appendix 1 presents the measures for cluster representatives.

5. RESULTS

This study examined the extent of communication exchange amplification in the forum core networks of multiple MOOC courses. Using an ERGM approach, we modelled the baseline density of networks, direct reciprocity of ties, and generalized exchange as the propensity of *simmelianties* network configuration to cluster. We have also controlled for the level of activity distinguishing between individuals with low, moderate and high posting behaviour. The research question inquired if any differences were observed between the measures of direct and generalized exchange in networks where staff implemented different facilitation strategies.

All fitted ERGMs have converged, were not degenerate, and showed acceptable goodness of fit. All models were fitted using the described network configurations. The two exceptions were course G where models did not converge with the effects of posting activity, and Course D where no super-posters were a part of the forum core. Table 4, 5 and 6 outline the results grouped together for courses with similar facilitation strategies.

First and foremost, our hypothesis that courses with no or low facilitation reflect the network structures typical for online public electronic groups was supported. In courses with low facilitation (Table 4) we observed higher propensity for direct exchange, and low or no propensity for generalized exchange. Course B, as a large course, had posters with high level of activity, and they were somewhat more likely to form ties than the posters with moderate participation, though the difference was not stark. Course G was very small, had no generalized exchange features, and the null model was only slightly improved by adding the direct reciprocity configuration.

In courses with moderate facilitation (Table 5), where teaching assistants, staff or community assistants were advancing the information flow, the dynamics had some similarities and differences with unmoderated settings. For the similarities, theta estimates for reciprocity in Courses D and F were higher than those for generalized exchange mirroring the dynamics of unmoderated courses. In three small courses (A, F and H) generalized exchange parameters were not significant, or even negative. Interestingly, the propensity of super-posters to form ties as compared to those posting moderately was more pronounced in courses with lower or no generalized exchange, and heightened reciprocity.

Table 4. ERGMs outputs for courses with low facilitation

Features/Courses	B	G
Density	**-4.21***** (0.03)	**-1.44***** (0.3)
	Structural Properties	
Reciprocity	**1.29***** (0.1)	1.84*** (0.3)
Generalized Exchange	**0.56***** (0.04)	--
	Main Effects	
Moderate Participation	**0.83***** (0.02)	--
High Participation	**1.66***** (0.05)	--
AIC Null	*38485*	*463*
AIC Final	*32926*	*442*

In contrast, course F with highest propensity for generalized exchange features to cluster, indicating higher network amplification, has least difference between the propensity for tie formation between those with moderate and high activity.

Table 5. ERGM outputs for courses with moderate facilitation

Features/ Courses	A	C	D	F	H
Density	**-1.66***** (0.06)	**-2.34***** (0.11)	**-2.93***** (0.08)	**-3.8***** (0.03)	**-3.06***** (0.16)
		Structural Properties			
Reci-procity	**1.82***** (0.3)	**1.35***** (0.35)	**0.09***** (0.22)	**1.57***** (0.1)	**1.45***** (0.26)
Generalized Exchange	**-0.55***** (0.13)	0.21. (0.13)	**0.46***** (0.09)	**0.66***** (0.06)	0.04 (0.09)
		Main Effects on Ties Formation			
Moderate Participation	**1.12***** (0.01)	**0.67***** (0.1)	0.068*** (0.06)	**0.85***** (0.04)	**0.59***** (0.1)
High Participation	**3.17***** (0.23)	**1.19***** (0.13)	--	**1.3***** (0.05)	**1.12***** (0.14)
AIC Null	*3084*	*1616*	*4045*	*24601*	*1451*
AIC Final	*2461*	*1334*	*3523*	*20451*	*1307*

These differences between a moderated Course A and unmoderated Course B can be interpreted as follows. In an unmoderated network information exchanges are random, distributed and decentralised. In moderated networks, moderators have much higher posting activity, sometimes 'dominating' in offering information to other learners. When this occurs, the generalized exchange features seem to be either low, negative or non-significant. By translating the log odds, we observe that the high participation posters in course A were 22 times more likely to form a tie as compared to low posters, while moderate posters were three times more likely. At the same time, in an unmoderated course B: posters with higher activity were six times as likely to make a post than those with low posting behavior, and moderate posters were three times more likely. We can speculate that in cases where super-posters do not over-dominate, activity of those with moderate posting behavior overtime grows, and these actors engage in more conversations with one another as well as interacting with super-posters, thereby resulting in an amplification of communication exchange.

Finally, network structures in courses with high facilitation differed from both moderated and unmoderated courses without instructor participation (Table 6). Overall, it appears that in courses with high facilitation, generalized exchange feature has a higher likelihood to occur than direct reciprocity. Such dynamics are demonstrated in courses E and J, where clustering of reciprocal triads are more characteristic of the network than person-to-person reciprocations. In a course I, however, direct reciprocity is still more characteristic of the network, which we interpret as the lack of amplification within the information exchange. Again, similar to the pattern observed with the moderately facilitated courses, in course I, high posting individuals are much more likely to form ties. In fact, by converting log odds, we found that a super poster in course I was eleven times more likely to form network ties than those with a low level of forum activity. In courses E and J the likelihood for super-posters to form ties was three times more than participants with a low posting behavior. To note, the individual number of posts by super-posters in course I was actually lower by count than that of super-posters in the other two courses.

Table 6. ERGM outputs for courses with high facilitation

Features /Courses	E	I	J
Density	**-3.69***** (0.02)	**-4*****(0.03)	**-3.21*****(0.04)
	Structural Properties		
Reciprocity	**0.93***** (0.17)	**1.91*****(0.15)	**0.65***(0.27)
Generalized Exchange	**1.13***** (0.08)	**0.31*****(0.06)	**1.05*****(0.13)
	Main Effects		
Moderate Participation	**0.53***** (0.023)	**1.33*****(0.04)	**0.73*****(0.03)
High Participation	0.98*** (0.05)	**2.41*****(0.07)	**1.32*****(0.06)
AIC Null	*27098*	*18244*	*18450*
AIC Final	*22005*	*13812*	*14347*

To extrapolate, the network dynamics is opposite between courses with high facilitation and low or no facilitation. Direct reciprocity of knowledge exchange seems to be inherent in identity-based communities, such as forum core. Overtime and with facilitative efforts network structure starts being defined by core and periphery: with both random reciprocation and low clustering of generalized exchange, likely at the core of the group. This clustering may also be interpreted as clique formation, or polarization of power and access that takes place as the network shifts from distributed to amplified. Facilitated forum core networks appear to be characterized by higher degree of clustering of reciprocated triads, and lower level of random reciprocation. This could mean that random direct exchanges within the group decrease, as information flow gets to amplify across more and more members, shifting from cliquish core and distributed periphery to an amplified interconnected cluster. We also observed that super-posting activity does seem to be associated with lower generalized exchange. In other words, a moderator may be taking over what somebody else could address by offering her services too much or too fast, therefore not allowing other members to indirectly reciprocate to the group.

To conclude, according to ERGM results, network features modelled to gain insight into network amplification were useful in highlighting differences between the courses with different facilitation strategies. More specifically, direct reciprocation was characteristic of non-facilitated groups, while generalized exchange was more prominent in highly facilitated online communities with instructor's involvement. Finally, super posting activity was less pronounced in networks with higher generalized exchange, and more pronounced in networks with higher direct reciprocity.

6. DISCUSSION

The aim of the current study was to establish potential indicators for evaluating social activity in MOOCs, particularly in relation to forum facilitation. Thus, in lieu of measuring affective individual perceptions as indicators of socio-emotional processes, the extent of amplification in communication exchanges between MOOC learners was examined. A sub-group of MOOC forum posters was conceptualized as an identity-based community. Building on the prior research on the properties of network formation in social exchange networks, we modelled ten forum core networks using network configurations of direct reciprocity and generalized exchange, while controlling for the overall level of individual activity.

Results support our hypotheses about the nature of the forum core in MOOCs, and suggest that the chosen indicators of forum core networks may be useful in evaluating the effects of facilitation. That is, courses with varying levels of moderation differ in the likelihood of direct and generalized exchanges that take place. In non-facilitated courses, the dynamics of the networks was similar to electronic networks of practice, and largely characterized by direct reciprocity. In highly moderated networks, generalized exchange configurations were more prominent than direct reciprocity. That is, these networks were amplified by the propensity of reciprocated ties in triads to cluster. Furthermore, a higher degree of generalized exchange seemed to take place in networks where activity of super-posting participants was less pronounced as compared to those posting moderately or at low levels.

The main contribution of the study is that its indicators of group-level social activities differentiated between the network structures in forums with different facilitation strategies. Our results demonstrated that moderating the forum *per se* is insufficient for the effective evolution of participation. Courses with teaching assistants and staff demonstrate different patterns than those with instructor involvement. In courses with moderate facilitation interventions generalized exchange features have lower likelihood to occur than in highly facilitated forums with instructor's involvement. These findings re-iterate the importance of teacher's social presence and its impact on the level of engagement in open online communities.

It is not surprising to learn that an instructor's presence motivates learners to participate - even in large scale courses. This finding has been widely discussed and supported in formal online education research. This study offers additional insights into the roles, played by the staff, teaching and community assistants. MOOC research claimed their importance in supporting the spirit of the forums. Our findings further demonstrate that moderated forums appear to transition from their baseline properties, as the structure shifts from distributed to more hierarchical network.

The results of the present study also support Huang et al.'s conclusion that super-posters do not drown out the silent majority. The observed interplay between the prominence of moderator involvement and the development of generalized exchange features suggest that the health of the forum may still be affected. In courses where super-posting activity is more distinct than that of moderate posters, features of generalized exchanged are not more pronounced than direct reciprocity features. Such dynamics may indicate that super-posters are limiting engagement opportunities for learners who are moderately active. In other words, if there is a moderator who always replies, and no teacher to stimulate the dialogue, then the community has less reason to engage with one another. It seems that super-posters, who are staff and teaching assistants at early stages, help the network to develop, but they are not always skillful at decreasing their activity at an optimal time.

This study's analysis yielded many promising results. However, further work is still required. Much of the interpretation of this work and the associated implications are at this point speculative. Future work should seek to account for the temporal aspect of how the network unfolds and interacts, as well as include forum core interactions with the rest of the forum posters in MOOCs. An extrapolation of the findings relates to the time based dynamics of network formation. We suggest that facilitation shifts the forum core network structure from a distributed to cliquish to egalitarian structure. Here an egalitarian structure is the product of combined direct and indirect reciprocity. Forum core is in its turn situated within the network exchanges of intermittent posters, and the relationship of forum core to those posters is based on the features of indirect reciprocity. That is, if a forum core member gets reciprocated by a staff member at an early stage, she would be more likely to offer answer to an intermittent poster later on. In other words, measuring the extent of network amplification over time offers partial explanation of the nature of social exchanges, while indirect reciprocation between the forum core and the other poster explains the remaining dynamics. These speculations form the basis of our future research.

Several direct implications stem from this study's analyses. First, empirical validation of forum core networks as identity-based communities is valuable as MOOC forum research can build on the established research agenda both in social science and in studies of human-computer interaction. For instance, Ren et al. [43] developed a set of design principles for the facilitation of identity-based communities, and these could be applied in MOOCs. Yet, as the dataset of courses was limited, more diverse sets of courses need to be analyzed to see if the insights demonstrate consistent patterns across disciplines, student cohorts and class sizes.

While this study did not include text mining and analysis of learner posts, identity-based online communities have been researched using text analysis along with SNA. Particularly [14] suggested that measures of entropy capture the lack of topical diversity in identity-based communities, what they refer to as 'topical groups', and the divergence of discussion themes in what they refer to as 'social groups'. Furthermore, granular text mining for the concepts related to power and authoritativeness [e.g. 23] may offer insights as to where the shifts in network structure are reflected in the text features of the various learner posts. On a final note, the indicators proposed from this work should be validated in formal online education environments, to assess for comparability between formal and open online settings.

7. ACKNOWLEDGMENTS

Our thanks to TU DELFT and Open Online Learning at DelftX.

8. REFERENCES

[1] Bell, F. (2011). Connectivism: Its place in theory-informed research and innovation in technology-enabled learning. The International Review of Research in Open and Distance Learning, 12(3).

[2] Brandes, U., Robins, G., McCranie, A. and Wasserman, S. (2013). What is network science? Network Science, 1(1), 1–15. http://dx.doi.org/10.1017/nws.2013.2

[3] Brinton, C.G., Chiang, M., Jain, S., Lam, H., Liu, Z. and Wong, F.M.F. (2013). Learning about social learning in MOOCs: From statistical analysis to generative model. IEEE Transactions on Learning Technologies, 7(4). http://dx.doi.org/10.1109/TLT.2014.2337900

[4] Brown, R., Lynch, C.F., Eagle, M., Albert, J., Barnes, T., Baker, R., Bergner, Y. and McNamara, D. (2015). Good communities and bad communities: Does membership affect performance? *Proceedings of the 8th International Conference on Educational Data Mining*, 612–614.

[5] Brown, R., Lynch, C., Wang, Y., Eagle, M., Albert, J., Barnes, T., ... McNamara, D. (2015). Communities of performance & communities of preference. In *CEUR Workshop Proceedings*. (Vol. 1446). CEUR-WS.

[6] Conole, G. (2013). *Designing for learning in an open world*. New York: Springer.

[7] De Laat, M., Lally, V., Lipponen, L. and Simons, R.-J. (2007). Online teaching in networked learning communities: A multi-method approach to studying the role of the teacher. *Instructional Science, 35*(3), 257–286. http://dx.doi.org/10.1007/s11251-006-9007-0

[8] Dowell, N. M., Skrypnyk, O., Joksimović, S., Graesser, A. C., Dawson, S., Gašević, S., … Kovanović, V. (2015). Modeling learners' social centrality and performance through language and discourse. In C. Romero & M. Pechenizkiy (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 250–257). International Educational Data Mining Society.

[9] Ferguson, R., Shum, S.B. (2012). Social learning analytics: five approaches. *Proceedings of the 2nd international conference on learning analytics and knowledge*, 23–33. http://dx.doi.org/10.1145/2330601.2330616

[10] Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education*, *23*, 18-26. http://dx.doi.org/10.1016/j.iheduc.2014.05.004

[11] Gillani, N., Yasseri, T., Eynon, R., & Hjorth, I. (2014). Structural limitations of learning in a crowd: communication vulnerability and information diffusion in MOOCs. *arXiv preprint arXiv*:1411.3662 [physics.soc-ph]. http://dx.doi.org/10.1038/srep06447

[12] Goldberg, L. R., Bell, E., King, C., O'Mara, C., McInerney, F., Robinson, A., & Vickers, J. (2015). Relationship between participants' level of education and engagement in their completion of the Understanding Dementia Massive Open Online Course. *BMC medical education, 15(1)*, 1. http://dx.doi.org/ 10.1186/s12909-015-0344-z

[13] Goodyear, P., Salmon, G., Spector, J. M., Steeples, C., & Tickner, S. (2001). Competences for online teaching: A special report. *Educational Technology Research and Development,* 49(1), 65-72. http://dx.doi.org/ 10.1007/BF02504508

[14] Grabowicz, P. A., Aiello, L. M., Eguíluz, V. M., & Jaimes, A. (2013). Distinguishing topical and social groups based on common identity and bond theory. In *Proceedings of the sixth ACM international conference on Web search and data*

mining, 627-636. http://dx.doi.org/10.1145/2433396.2433475

[15] Gruzd, A., & Haythornthwaite, C. (2013). Enabling community through social media. *Journal of medical Internet research,* 15(10), e248.

[16] Gruzd, A. & Haythornthwaite, C. (2008). Automated discovery and analysis of social networks from threaded discussions. Paper presented at the International Network of Social Network Analysis, St. Pete Beach, FL, USA.

[17] Handcock, M., Hunter, D., Butts, C., Goodreau, S., Krivistky, P. and Morris, M. 2015. *ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks*. The Statnet Project. http://www.statnet.org.

[18] Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2003). statnet: Software tools for the Statistical Modeling of Network Data. Seattle, WA. Version, 2.

[19] Harris, J. K. (2013). *An introduction to exponential random graph modeling* (Vol. 173). Sage Publications.

[20] Haythornthwaite, C. and De Laat, M. (2012). Social network informed design for learning with educational technology. In A.D. Olofsson, J.O. Lindberg, K. Klinger, C. Shearer (Eds.), *Informed Design of Educational Technologies in Higher Education: Enhanced Learning and Teaching* (pp.352–374). IGI Global.

[21] Hecking, T., Chounta, I.-A. and Hoppe, H.U. (2016). Investigating social and semantic user roles in MOOC discussion forums. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 198–207. http://dx.doi.org/10.1145/2883851.2883924

[22] Hecking, T., Hoppe, H.U. and Harrer, A. (2015). Uncovering the Structure of Knowledge Exchange in a MOOC Discussion Forum. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, 1614–1615. http://dx.doi.org/10.1145/2808797.2809359

[23] Howley, I., Mayfield, E. and Rosé, C.P. (2011). Missing something? Authority in collaborative learning. In A. Dimitracopolou, C. O'Malley, D- Suthers, P-Reimann (Eds.), *Proceedings of the 9th Computer Supported Collaborative Learning Conference (*pp.336373).

[24] Huang, J., Dasgupta, A., Ghosh, A., Manning, J. and Sanders, M. (2014). Superposter behavior in MOOC forums. *Proceedings of the first ACM conference on Learning @ Scale*, 117–126.

[25] Jiang, S., Warschauer, M., Williams, A.E., O'Dowd, D. and Schenke, K. (2014). Predicting MOOC performance with week 1 behavior. *Proceedings of the 7th International Conference on Educational Data Mining*, 273–275.

[26] Joksimović, S., Dowell, N., Skrypnyk, O., Kovanović, V., Gašević, D., Dawson, S. and Graesser, A.C. (2015). How do you connect? Analysis of social capital accumulation in connectivist MOOCs. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, 64–68. http://dx.doi.org/10.1145/2723576.2723604

[27] Joksimović, S., Manataki, A., Gašević, D., Dawson, S., Kovanović, V. and de Kereki, I.F. (2016). Translating network position into performance: importance of centrality in different network configurations. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 314–323. http://dx.doi.org/10.1145/2883851.2883928

[28] Jones, C. (2004). Networks and learning: communities, practices and the metaphor of networks. *Research in Learning Technology, 12*(1).

[29] J Justice, T., & Jamieson, D. W. (2012). *The facilitator's fieldbook*. AMACOM Div American Mgmt Assn.

[30] Kellogg, S., Booth, S. and Oliver, K. (2014). A social network perspective on peer supported learning in MOOCs for educators. *The International Review of Research in Open and Distributed Learning*, *15*(5).

[31] Kizilcec, R.F., Piech, C. and Schneider, E. (2013). Deconstructing disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses. *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13*.

[32] Krackhardt, D. (1998). Simmelian ties: Super strong and sticky. In R. Kramer, M. Neale (Eds.), *Power and influence in organizations* (pp. 2138). SAGE Publications: USA.

[33] Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. Computers in human behavior, 19(3), 335-353. http://dx.doi.org/10.1016/S0747-5632(02)00057-2

[34] Kreijns, K., Kirschner, P. A., Jochems, W., & Van Buuren, H. (2011). Measuring perceived social presence in distributed learning groups. *Education and Information Technologies,* 16(4), 365-381. http://dx.doi.org/10.1007/s10639-010-9135-7

[35] Laumann, E., Marsden, P. and Prensky, D. (1983). The boundary specification problem in network analysis. In L.Freeman, D. White (Eds.), *Research methods in social network analysis* (pp.6186). Transaction Publishers: London, UK.

[36] Lusher, D., Koskinen, J. and Robins, G. (2012). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge University Press: USA.

[37] Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of exponential-family random graph models: terms and computational aspects. *Journal of statistical software*, 24(4), 1548 7660.

[38] Poquet, O. (2016, September). Needle in a haystack: Analysis of social processes in MOOCs. Presentation at the EASS HDR Forum, School of Education, University of South Australia.

[39] Poquet, O., Kovanovic, V., de Vries, P., Hennis, T., Joksimovic, S., Gasevic, D. and Dawson, S. (n.d.) *Social presence in Massive Open Online Courses*. Manuscript submitted for review.

[40] Poquet, O. and Dawson, S. (2016). Untangling MOOC learner networks. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 208–212. http://dx.doi.org/10.1145/2883851.2883919

[41] Postmes, T., Spears, R., & Lea, M. (1999). Social identity, normative content, and "deindividuation" in computer-mediated groups. In R. Spears, B. Doosjie & N. Ellemans (Eds.), *Social identity: Context, commitment, content* (pp. 164183). Oxford: Blackwell.

[42] Ren, Y., Harper, F. M., Drenner, S., Terveen, L. G., Kiesler, S. B., Riedl, J., & Kraut, R. E. (2012). Building Member

Attachment in Online Communities: Applying Theories of Group Identity and Interpersonal Bonds. *Mis Quarterly, 36*(3), 841-864.

[43] Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization studies, 28*(3), 377-408. http://dx.doi.org/10.1177/0170840607076007

[44] Robins, G., Pattison, P., & Elliott, P. (2001). Network models for social selection processes. *Psychometrika*, *66*(161). http://dx.doi.org/10.1007/BF02294834

[45] Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2007). Assessing social presence in asynchronous text-based computer conferencing. *International Journal of E-Learning & Distance Education, 14*(2), 50-71.

[46] Rovai, A. P. (2002). Building sense of community at a distance. *The International Review of Research in Open and Distributed Learning, 3*(1).

[47] Sassenberg, K. (2002). Common bond and common identity groups on the Internet: Attachment and normative behavior in on-topic and off-topic chats. Group Dynamics: Theory, Research, and Practice, 6(1), 27. http://dx.doi.org/10.1037/1089-2699.6.1.27

[48] Siemens, G. (2004, December). Connectivism: A learning theory for the digital age. [Web log post]. Retrieved from http://www.elearnspace.org/Articles/connectivism.htm

[49] Siemens, G. (2008). *Learning and Knowing in Networks: Changing Roles for Educators and Designers*. Presented at ITFORUM.

[50] Skrypnyk, O., Joksimovic, S., Kovanovic, V., Gasevic, D. and Dawson, S. 2014. Roles of course facilitators, learners, and technology in the flow of information of a cMOOC. *The International Review of Research in Open and Distributed Learning, 16*(3).

[51] Utz, S. (2003). Social identification and interpersonal attraction in MUDs. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie*, 62(2), 91. http://dx.doi.org/10.1024//1421-0185.62.2.91

[52] Vu, D., Pattison, P. and Robins, G. (2015). Relational event models for social learning in MOOCs. *Social Networks*, 43, 121–135. http://dx.doi.org/10.1016/j.socnet.2015.05.001

[53] Wang, X., Wen, M. and Rosé, C.P. (2016). Towards triggering higher-order thinking behaviors in MOOCs. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 398–407. http://dx.doi.org/10.1145/2883851.2883964

[54] Wasko, M., Teigland, R. and Faraj, S. (2009). The provision of online public goods: Examining social structure in an electronic network of practice. *Decision Support Systems*, *47*(3), 254–265. http://dx.doi.org/10.1016/j.dss.2009.02.012

[55] Yang, D., Sinha, T., Adamson, D. and Rose, C. (2013). “Turn on, Tune in, Drop out”: Anticipating student dropouts in Massive Open Online Courses. Presented at the NIPS Workshop on Data Driven Education, Lake Tahoe, Nevada, USA.

Appendix 1.

Table A1. Results of k-means clustering of in-reach and out-reach measures derived from the entire MOOC forum network in each of the courses. Attributes for clusters were used to control for activity in forum core ERGM. In most courses clusters were interpreted as posters with high, moderate and low posting activity (except for C)

Course	A	B	C	D	E	F	G	H	I	J
Cluster 3 - High Posting Activity										
Posters, N	**3**	**6**	**50**	**1**	**4**	**7**	**7**	**11**	**3**	**4**
In-reach Centroid	13525	6045	252	262089	50415	18797	601	641	14630	69788
Out-reach Centroid	18892	7808	835	178378	58515	22576	917	672	12994	81533
Cluster 2 - Moderate Posting Activity										
Posters, N	**65**	**273**	**75**	**398**	**459**	**40**	**58**	**61**	**38**	**41**
In-reach Centroid	2612	655	578	2748	2337	4484	214	182	2054	16114
Out-reach Centroid	2536	710	221	2900	2375	6116	212	223	2864	21084
Cluster 1 - Low Posting Activity										
Posters, N	**503**	**2998**	**381**	**856**	**1297**	**4082**	**262**	**849**	**1994**	**1317**
In-reach Centroid	475	32	20	144	116	170	15	14	65	743
Out-reach Centroid	463	35	20	142	91	159	19	12	58	619