

# Is Participating in MOOC Forums Important for Students? A Data-driven Study from the Perspective of the Supernetwork

Chaocheng He, Panhao Ma, Lusha Zhou, Jiang Wu<sup>†</sup>

School of Information Management, Center for E-commerce Research and Development,  
Wuhan University, Luojia Mountain, Wuhan 430072, China

Citation: Chaocheng He, Panhao Ma, Lusha Zhou, Jiang Wu (2018). Is Participating in MOOC Forums Important for Students? A Data-driven Study from the Perspective of the Supernetwork.

Vol. 3 No. 2, 2018

pp 62–77

DOI: 10.2478/jdis-2018-0009

Received: Dec. 16, 2017

Revised: Feb. 5, 2018

Accepted: Mar. 22, 2018

## Abstract

**Purpose:** Compared with traditional course materials used in the classroom, the massive open online course (MOOC) forum that delivers unlimited learning content to students has various advantages. Yet MOOC has also received criticism recently, notably the problem of extremely low participation rates in its discussion forums. This study aims to explore the correlation between forum activity and student course grade in MOOC, and identify more accurately the forum activity levels of participants and the quality of threads in MOOC.

**Design/Methodology/Approach:** We crawled students' tests, final exams, exercises, discussions performance data and total scores from a course in Chinese College MOOC from May 2014 to August 2014. And we use the data to analyze the correlation between Forum Participation and Course Performance based on nonparametric tests as well as multiple linear regressions with the software of R. The study provides definitions and algorithms of super degrees based on the supernetwork model to help find high-quality threads and active participants.

**Findings:** A positive correlation between forum activity and course grade is found in this study. Students who participate in the forum have better performance than those who do not. Using the definitions and algorithms of super degrees in the supernetwork, forum activity levels of participants as well as the quality of threads they employ are identified.

**Research limitation:** Only limited representative forum participants and threads are used to analyze the activity level and significance of the MOOC forum. Also, the study only investigates one Chinese course on information retrieval. More data and more data sources could be helpful in better understanding the MOOC forum phenomenon.

**Practical implications:** As super degrees can reveal more latent information and recognize high-quality threads as well as active participants, these parameters can be used to assess needs to improve forum settings and alleviate the problem of low forum participation. The proposed super degrees can be applied in social network domains for further research.



**Originality/Value:** Definitions and algorithms of super degrees are provided and used for forum analysis. Super degrees can be applied to find high-quality threads and active participants, which is beneficial to guide students to participate in these high-quality threads and have a better understanding of knowledge MOOC provides.

**Keywords** MOOC; Participate; Forum; Supernetwork; Super degrees

## 1 Introduction

With the development of Web 2.0 technology and trends in education globalization, the massive open online course (MOOC) has rapidly become an influential educational tool. Enabling students from all over the world to access free education provided by elite universities (Daniel, 2012), and offering the meeting point for communities of people that share common interests (McAuley et al., 2010), MOOC is not only a great innovation in internet applications, but also a revolution in education. Following the success of Coursera and edX in the United States, many initiatives have been created in China since 2013, such as Xuetaangx, CNMOOC, and Chinese College MOOC. The use of MOOC, however, has recently received criticism from teachers and scholars.

One of the most frequent criticisms of MOOC relates to low enrollment rates in forum discussions. Breslow et al. (2013) reported that only three percent of all students enrolled in edX's first MOOC participated in the discussion forum. Duke University reported that about seven percent of participants in its first MOOC contributed to the forum (Belanger & Thornton, 2013). The University of Edinburgh had slightly higher participation numbers, where an average of 15 percent of participants posted in forums of its first six MOOCs (MOOCs@Edinburgh Group, 2013). Zhang et al. (2015) found that approximately ten percent of learners complete MOOCs, where the absence of peer and professor support are seen to contribute to retention issues.

Several scholars have investigated the reasons why these learners do not participate in these forums. Alario-Hoyos et al. (2014) found that some people with negative intentions take advantage of the open nature of MOOC, where they post off-topic comments, personal promotions, or spam. The phenomenon of information overload is also very serious, since there are too many repeated and irrelevant posts in forums. Although there is an abundance of data available in MOOCs, it is often difficult to obtain useful, relevant information when needed (Edmunds & Morris, 2000). Teachers and students cannot easily find valuable information.

It is important to note, however, that participating in these forums has been beneficial for learners. Students can exchange ideas or critiques, communicate queries, and benefit from co-construction of knowledge within MOOC discussion



forums without having to pay for the privilege (Hull & Saxon, 2009). This level of discussion and interaction is helpful for knowledge acquisition and improving the effects of learning (Zhao 2014). With discussion forum settings and certain assessment criteria in place to help learners find high-quality threads and active participants, the benefits of MOOC increase.

This study proposes definitions and algorithms of super degrees (including super-node degree and super-edge degree) in the supernetwork (or supernet), which is an “Internet Protocol (IP) network that is formed, for routing purposes, from the combination of two or more networks (or subnets) into a larger network” (Wikipedia, <https://en.wikipedia.org/wiki/Supernetwork>). Based on the perspective of the supernetwork employing super-degree parameters, the active participants in MOOC forums as well as the superior threads released by forum participants can be identified. In this way the problem of information overload can be alleviated, and more students will participate in the forum.

This paper aims to investigate how MOOC forums are truly important and correlate with students’ improved academic performance. It also confirms the advantages of super degrees algorithms, which partly fill a gap of supernetwork parameters, thus we can better identify forum activity levels of participants and the quality of threads in MOOC, and more students will participate in the forum.

The rest of this paper is organized as follows. Section 2 offers a more detailed review of MOOC forum studies. Section 3 details nonparametric tests and multiple linear regressions that are used to find the relation between forum participation activities and course scores. In section 4, from the perspective of the supernetwork, super-node degree and super-edge degree are defined and employed with real course data to do forum analyses. Section 5 draws conclusions and introduces visions for future researches emerging from this work.

## 2 Related Work

Along with the two MOOC networks, the people network and knowledge network, there are two basic MOOC curriculum modes used by students and teachers: cMOOC and xMOOC. In 2008, George Siemens and Stephen Dawnes created the first MOOC: Connectivism and Connective Knowledge Online Courses (Siemens, 2014) which is later called cMOOC as it is based on the connectivism learning theory. cMOOC highlights the important role that social and cultural context plays in how and where learning occurs. Connectivism sees knowledge as “a network and learning as a process of pattern recognition” (Wikipedia, <https://en.wikipedia.org/wiki/Connectivism>). Learning in this case happens not only in the individual context but within and across large networks, including organizations and big data. Focus



here is on connecting specialized information sets that rely on user-generated content, wherein connections among participants serve as keys for the course to advance (Siemens, 2014). The MOOC forum thus offers a uniquely broad platform for participants to discuss what they learn and to build creative social connections, which also aligns with the connectivism theory (Downes, 2010). With the representatives of Coursera, Udacity, edX which were fast growing in 2012, xMOOC is quite different from cMOOC. xMOOC derives from behaviorism, replicating the traditional educational model of knowledge transfer from teachers to students and emphasizes courses, exercises, and tests based on behaviorism learning theory. Combining aspects of philosophy, methodology, and psychological theory, behaviorism assumes that all human and animal behaviors are either reflexes produced by a response to stimuli in the environment, or a result of an individual's history, such as reinforcement and punishment. Although behaviorists "generally accept the important role of inheritance in determining behavior, they focus primarily on environmental factors" (Wikipedia, <https://en.wikipedia.org/wiki/Behaviorism>).

Although there has been extensive researches on MOOC in recent years, limited studies address forum discussion issues. Several papers (e.g. Belanger & Thornton, 2013; Breslow et al., 2013; Manning & Sanders, 2013) mentioned that most forum participation rates are about five percent, a number that is far overshadowed by the high registration rates in MOOC. As Engle et al. (2015) summarized, the students who completed most of the course assessments (Kizilcec, Piech, & Schneider, 2013) or earned a certificate (Ho et al., 2014) were more likely to be discussion forum participants. And in edX's first MOOC offering, 52 percent of certificate earners were active on the forum. In a MOOC addressing business strategy, students who posted on the forum were more likely to complete the course successfully (Gillani & Eynon, 2014). Huang et al. (2014) studied a special cohort "superposter," defined as the highest-volume forum contributors. They replied faster, received more up-votes, and obtained higher grades than the average forum participants. From these studies, it is not hard to find that there is a distinct correlation between forum participation and higher grades.

Other researches addressing the MOOC forum include Yang et al. (2014) explored how peer relations in MOOC forums influence student dropout rates. Rossi et al. (2014) and Stump et al. (2013) both analyzed forum content and classified discussion forum posts. Using standard social network analytic techniques, Yang et al. (2013) explored factors related to student behavior and social positioning within discussion forums. Similarly, Kellogg et al. (2014) studied the social networks in the forum. Anderson et al. (2014) investigated how forum participation relates to other parts of the course, where they found that badges can serve as incentives for engagement in a MOOC, including its forum.



### 3 Analysis on the impact of Forum Participation on Course Performance

Chinese College MOOC is one of the biggest MOOC platforms in China. Like other MOOC platforms, it provides instructional videos, exercises, tests, and interactive forums for students. In this section, the course “Information Retrieval” in Chinese College MOOC is used as the research subject. Course data is crawled and analyzed through nonparametric tests as well as multiple linear regressions to investigate the influence of participating in discussion forums on student academic performance.

#### 3.1 Data source

To ascertain the influence of forum participation on student grades, we first obtain students’ course information, including userID, scores for tests, exercises, final exams, and discussions, as well as total scores and discussion activities (e.g. the number of comments, replies, and posts). The Information Retrieval course examined in this study opened in May 2014 with 3,231 registrants.

In this course, the full marks for test scores, exercises scores, final exam scores and the full score for discussions are 113, 30, 30, and 30 respectively, accounting for 30 percent, 10 percent, 30 percent, and 30 percent of the total score. Especially, if the number of student posts, comments, or replies is higher than three, the full discussion scores can be identified (Table 1).

Table 1. Descriptive statistics of student scores and forum activities.

	N	Minimum	Maximum	Mean	Std. Deviation
Tests/113.0	1,917	0	113	69.23	38.98
Exercises/30.0	2,625	3	30	15.30	7.21
Exams/30.0	1,264	0	30	28.38	2.67
Discussions/30.0	3,231	0	30	22.10	13.21
Posts	804	0	15	0.57	1.46
Comments	804	0	46	0.43	2.29
Replies	804	0	116	3.59	8.02
Total Score/100	3,231	0	99.8	48.15	34.97

#### 3.2 Nonparametric tests

Students in this study are divided into a discussion group and non-discussion group according to whether they participate in the MOOC discussion forum, where differences in scores between the two groups include test statistics and the nonparametric analysis results (Table 2).



Table 2. Test statistics.

	Test Score	Exercises Score	Exam Score	Total Score
Z	-4.439	-5.06	-5.983	-3.918
P-value	.000	.000	.000	.000

Table 3. Results of nonparametric tests.

	Group	N	Mean Rank	Mean score
Tests	Non-discussion	1,343	798.94	58.97
	Discussion	573	1,332.49	94.31
Exercises	Non-discussion	2,056	1,138.56	13.5
	Discussion	568	1,942.11	21.77
Exams	Non-discussion	755	591.66	28.02
	Discussion	508	691.96	28.87
Scores	Non-discussion	2,599	1,400.70	40.14
	Discussion	631	2,500.22	81.05

P-values in Table 2 indicate that the test scores, exercises scores, exam scores, and total scores of the two groups have significant differences. And we can infer from the mean rank in Table 3 that grade of the discussion group is better than the non-discussion group, so participating in a discussion forum can truly improve student achievement in MOOC.

### 3.3 Model estimation

Besides the nonparametric tests, multiple linear regressions are used to investigate the relationship between forum activities and scores. Taking the numbers of comments, posts, and replies as independent variables, the dependent variable is the total score weeding out the discussion part that students participating discussion forum get in the MOOC course. Comment refers to other students' comment on one student, while reply is the one student's reply to others, and post refers to the students' posting topics in the forum. The proposed regression model is as follows:

$$\text{Score} = \text{Constant} + \beta_1 \text{Comment} + \beta_2 \text{Reply} + \beta_3 \text{Post} + \epsilon.$$

The numbers of comments, replies, and posts are found to reflect MOOC users' involvement in the forum, where students can exchange ideas and benefit from co-construction of knowledge by their participation. We therefore suggest that forum participation will have a significantly positive influence on student grades. To verify our assumption, we do multiple linear regression analysis (Table 4).



Table 4. Regression analysis results.

	Coefficient	P-value
Comment	0.879	.004
Reply	3.937	.000
Post	1.776	.000
(Constant)	-.900	.242
Adjusted R <sup>2</sup>		0.657
P-value		0.000

The regression gets good fitting results (adjusted  $R^2 = 0.657$ ,  $p = 0.000$ ). Table 4 indicates that comment ( $\beta_1 = 0.879$ ,  $p = 0.004$ ), reply ( $\beta_2 = 3.937$ ,  $p = 0.000$ ) and post ( $\beta_3 = 1.776$ ,  $p = 0.000$ ) all have significantly positive effects on student grades. Participating in forums is found to contribute to communication, where discussions with others can help deepen the understanding of the course content. This means that the MOOC forums can reinforce learning, help students to better master knowledge, and thus play an important role in enhancing grades.

Based on above research, we find that forum participation activities do have a positive influence on students' scores, as discussion group members (people who participate in the forum) get higher scores than those in the non-discussion group.

## 4 Forum Analysis Based on Supernetworks

While study outcomes confirm that participation in forums is beneficial to students' performance in MOOC, A new assessment criteria is thus proposed, where participants' activity levels and the importance of threads in discussion forums are analyzed based on supernetwork. The algorithms and significance of the super degrees is also used to analyze participants and threads of one discussion forum in MOOC.

### 4.1 What is a supernetwork?

Yosef Sheffi (1985) first put forward the concept of the "supernetwork" model as a suitable technique for integrated system modeling. Nagurney et al. (2002) referred to the supernetwork as being above and beyond existing networks. Single networks cannot completely depict characteristics of the real world and relations between networks, whereas a supernetwork can describe and express network information more comprehensively and clearly.

In research on supernetwork, the use of the hypergraph is frequently employed. In mathematics, a hypergraph is a "generalization of a graph in which an edge can join any number of vertices... a pair that is a set of elements called nodes or vertices, and is a set of non-empty subsets of called hyperedges or edges" (Wikipedia,



<https://en.wikipedia.org/wiki/Hypergraph>). The defining formula of the hypergraph is as follows (Berge & Minieka, 1973):

Assuming  $V = \{v_1, v_2, \dots, v_n\}$  is a finite set, if

- $E_i \neq \emptyset \quad (i = 1, 2, \dots, m)$ ,
- $\bigcup_{i=1}^m E_i = V$ ,

then the binary relation  $H = (V, E)$  is called the hypergraph.  $V = \{v_1, v_2, \dots, v_n\}$  is called the hypergraph vortex.  $E = \{e_1, e_2, \dots, e_m\}$  is the edge set of hypergraph, and set  $E_i = \{v_{i_1}, v_{i_2}, \dots, v_{i_j}\}$ ,  $(i = 1, 2, \dots, m)$  is called hypergraph edge, which is the super-edge, exhibited in the following formula and Figure 1):

$$V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$$

$$E = \{e_1 = \{v_1, v_2, v_3\}, e_2 = \{v_1, v_4\}, e_3 = \{v_2, v_3\}, e_4 = \{v_3, v_5, v_6\}, e_5 = \{v_4, v_7\}\}$$

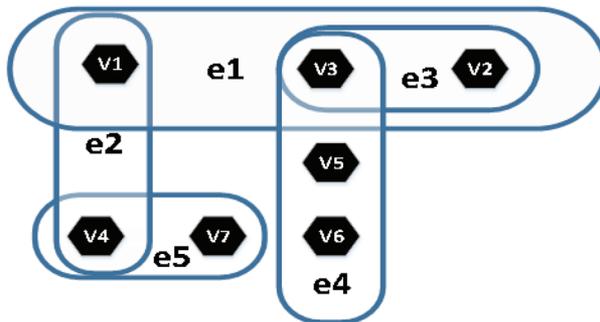


Figure 1. Example of a Hypergraph.

In MOOC, there are two networks, the people network and knowledge network. People of all ages, with different education backgrounds coming from different countries, consist of the people network. Different fields of knowledge offered by MOOC platforms constitute the network. The people and knowledge networks interlace and build connections with each other, so that they constitute multi-style, multi-level, and multi-dimensional supernetworks. In this complicated supernetwork, we employ some parameters to depict it, which can describe the forum's activity levels more clearly and help us better understand active levels of participants as well as the degree of the threads' significance.

## 4.2 Super degrees of the supernetwork

Studies about supernetworks are still in the development stage. Although the concept of the supernetwork is defined, the parameters to depict it still lacks unified and specific definitions. Taking properties of the super-node and super-edge into



consideration, we define the concepts and algorithms of super-node degree and super-edge degree (collectively called “super degrees”) based on the hypergraph, as follows:

**Node degree:** In the hypergraph, node degree of super-node  $v_i$  is defined as the sum of super-edges, including  $v_i$ , noted as  $d_{e_j}(v_i)$ ;

**Edge degree of super edge:** In the hypergraph, the edge degree of the super-edge  $e_j$  is defined as the sum of super-nodes belonging to  $e_j$ , noted as  $d_{v_i}(e_j)$ ;

**Super node degree:** In the hypergraph, the super-node degree of super-node  $v_i$  is noted as  $d_H(v_i)$ , where the formula is:

$$d_H(v_i) = d_{e_j}(v_i) \times \frac{\sum_i d_{v_i}(e_j)}{\sum_j d_{v_i}(e_j)} \quad (1)$$

(Note,  $\sum_i d_{v_i}(e_j)$  means the sum of the edge degree of super-edges including  $v_i$ , and  $\sum_j d_{v_i}(e_j)$  means the sum of the edge degree of all super-edges);

**Super edge degree:** In the hypergraph, the super-edge degree of super-edge  $e_j$  is noted as  $d_H(e_j)$ , where the formula is:

$$d_H(e_j) = d_{v_i}(e_j) \times \frac{\sum_j d_{e_j}(v_i)}{\sum_i d_{e_j}(v_i)} \quad (2)$$

(Note,  $\sum_j d_{e_j}(v_i)$  means the sum of node degree of super-nodes belonging to  $e_j$ ,  $\sum_i d_{e_j}(v_i)$  means the sum of node degree of all super-nodes.

To better understand the definitions of these degrees, take Figure 1 for example. The node degrees of super-nodes  $v_1, v_2, v_3, v_4, v_5, v_6, v_7$  are 2, 2, 3, 2, 1, 1, and 1, respectively. The edge degrees of super-edges  $e_1, e_2, e_3, e_4, e_5$  are 3, 2, 2, 3, and 2, respectively. The super-node degrees of super-nodes  $v_1, v_2, v_3, v_4, v_5, v_6, v_7$  are 0.83, 0.83, 2.00, 0.67, 0.25, 0.25, and 0.17, respectively. The super-edge degrees of super-edges  $e_1, e_2, e_3, e_4, e_5$  are 1.75, 0.67, 0.83, 1.25, and 0.50, respectively. The functions of these parameters in MOOC are as follows:

- Node degree denotes the frequency in which a certain person participates in the forum threads. The larger the node degree value, the more active this participant is, and the better he or she facilitates knowledge sharing.
- Edge degree denotes the number of people participating in certain threads (representing knowledge levels). The larger the edge degree value, the more likely thread is attractive, and better it facilitates knowledge sharing.
- Super node degree denotes the activity level of a certain person participating in the forum, and her or his ability to facilitate knowledge sharing. According



to its definition, the super-node degree is of relative magnitude, different from the node degree. That is, if one person participates in threads with higher participation rates, the super-node degree of that person is larger compared with another with the same node degree. It is then more likely that those with larger audiences play a more significant role in the forum, because they can influence more students and contribute more to facilitating knowledge sharing.

- Super edge degree denotes a certain thread's significance or attraction in appealing to more forum participants. As the same argument, it is also relative in magnitude. If one thread attracts more people with higher participation rates, then compared with other threads with the same edge degree, it has a larger super-edge degree. Because it attracts people who are more active and influential, this thread can exert more effect on readers.

### 4.3 Super degrees analysis of one MOOC

This study uses a crawled dataset retrieved from the discussion forum of the course "Information Retrieval." There are many threads in the forum, and students can post to ask for help, to discuss the lectures, to share learning notes, and so on.

Among the data crawled, we pick up five threads (marked as threads 1 through 5) that not only have more participants and the number of posts but also have significant gradients, that of "Week 7–Interaction Assignment," "Week 4–Interaction Assignment," "Week 2–Interaction Assignment," "Week 1–Interaction Assignment" and "Information retrieval and information selection". The post numbers of these six threads are 106, 97, 85, 42, and 31, respectively. The weight of super-nodes and super-edges are not taken into consideration when calculating parameters of the supernetwork. We therefore adjust the numbers of people who post several times at the same threads. This means that if someone posts several times at the same thread, we only mark it as one time. Then the adjusted numbers of threads are respectively 99, 86, 78, 36, and 24.

We calculate the numbers of participants in these six threads, and pick up students who participate in more than one thread. The finding is that 3 people (e.g. "Viannn" username) participate in five threads; 7 people (e.g. "mooc15951364231" username) participate in four threads, and the rest participate in three or fewer threads. The majority of people only participate in one or two threads. Selecting 10 typical participants, Table 5 depicts their thread participation levels.

The relevant parameters of above participants and their threads is then calculated, where node degrees and super-node degrees are calculated (Table 6).

Node degrees and super-node degrees all indicate participants' MOOC forum activity levels. The larger the value is, the more active this person is. Nevertheless,



## Research Paper

Table 5. Participants and threads they participate in.

Username	Thread 1	Thread 2	Thread 3	Thread 4	Thread 5
Viannn	√	√	√	√	√
Winner	√	√	√	√	
Day@4	√	√	√		√
moooc15951364231	√	√		√	
Lin Wei ykt1123		√	√	√	
Red Fruit moooc4	√	√			
Chongqing WZZ	√		√		
AYmoooc	√			√	
Monogram	√				√
SUNSET		√			√

Table 6. Node degree and super-node degree of participants.

Username	Node degree	Super-node degree
Viannn	5	5
Winner	4	3.59
Day@4	4	3.31
moooc15951364231	3	2.17
Lin Wei ykt1123	3	1.86
Red Fruit moooc4	2	1.10
Chongqing WZZ	2	1.10
AYmoooc	2	0.90
Monogram	2	0.90
SUNSET	2	0.76

super-node degrees may be different even though people have the same node degree. For example, “Day@4” and “moooc15951364231” have the same node degree, but the former has a larger super-node degree. This means that he or she participates in much more threads with higher participation rates, and knowledge shared or opinions expressed may influence more people. The Super-node degree of “Viannn” and “Winner” rank top 2, which indicates that they have higher activity levels and thus a larger potential to acquire and share knowledge. But “Shadowfollower” et al. have relatively smaller values, indicating that they are less active in the forum and may exert a smaller influence on others.

Considering the entire forum, however, participants are far fewer in number than mere course registrants, where the numbers of threads that most people participate in are very small (one or two). Most students in the course do not participate in the forum at all. So the majority of people have smaller node degrees and super-node degrees than participants listed in Table 7. Edge degrees and super-edge degrees of the six threads are shown in Table 7.



Table 7. Edge degree and super-edge degree of threads.

Thread	Edge degree	Super-edge degree
Thread 1	8	6.62
Thread 2	8	6.90
Thread 3	5	3.10
Thread 4	5	2.93
Thread 5	3	1.14

Edge degree and super-edge degree both indicate the threads’ participation levels. The larger the value is, the more people the thread attracts. What’s more, the larger super-edge degree indicates that this thread could allow more influential people with higher participation rates to join in to facilitate knowledge sharing. Threads 1 to 5 represent the participation rates with different levels. Threads 1 and 2 have larger super-edge degrees than others, which means that they are more important to participants. Threads 3 and 4 have the same edge degree but different super-edge degrees, where Thread 3’s larger super-edge degree indicates that it attracts more people with higher participation rates and thus contributes more to knowledge sharing.

Nowadays, most MOOC platforms only offer the number of viewers and comments to students, where only a small percentage also provide numbers of up-votes, which represent positive (negative) votes for threads. But it is not enough to manifest the quality of threads. As some researchers said, the more we enable participants to know about their interaction information and quality of threads, the more they will be motivated to join in the forum. In the same vein, the more students depend on interaction information, the more actively they would participate in the forum discussion (Zhan et al., 2015). It is therefore important to improve forum settings and offer suitable assessments of student (and teacher) contributions as well as the quality of threads.

The study parameters of super-node degree and super-edge degree can therefore act as an assessment of activity levels of forum participants as well as degrees of thread significance. It is easy to understand that one thread may have many posts, but if most posts are off-topic or worthless, the number of posts cannot truly reflect the value of this thread. But one thread with a higher participation rate is more likely to be a valuable thread, where super-edge and super-node degrees can reveal this latent or hidden information.

Above all, the super degrees can be used in MOOC platforms to act as a kind of assessment, for it can partly express activity levels as well as the significance and quality of participants’ threads.



## 5 Conclusion and Future Work

Compared with traditional education taking place in actual (versus virtual) classrooms, MOOC lacks the interaction process and stimulation among students. We find there is a positive correlation between forum activity and course grade. In the forum, students can build study communities, discuss learning achievements, and encourage each other to excel in or complete the course. Nonetheless, MOOC forum participation rates are very low (between five and ten percent). This common phenomenon means that there is a need to improve the forum's data interpretation, as well as to emphasize the importance and advantages of participating in the forum for students.

Employing background data of one course "Information Retrieval" in Chinese College MOOC, we make nonparametric tests to show that people who participate in the discussion forum over-perform than non-discussion people in academic tests, exercises, and exams. We also make multiple linear regressions to verify forum activities such as comment, reply, and post behaviors, all of which have positive effects on student grades.

While participating MOOC forums improves study outcomes in terms of grades, many students choose not to join these forums. This study proposes new assessment criteria to help solve this problem from the perspective of the supernetwork. Through actively participating in the forum, people could build deeper connections with diverse information and large datasets, which contributes to spreading knowledge more widely, and sharing knowledge more quickly and efficiently. This tool can thus help students to improve their learning effects and subsequently their grades. Employing forum data of the course "Information Retrieval" in Chinese College MOOC together with super degrees parameters, we find that use of super degrees can help us better understand activity levels of participants as well as the quality of threads. We thus propose that these parameters can act as a kind of assessment to let students access more information that can be effective in meeting their learning goals. Then they would be more likely to participate in the MOOC forum.

Limitations of this study are the difficulty of data acquisition, and the specific data sample limited to one information retrieval course in China. A broader study sample and topic with students from other countries would be helpful in understanding the MOOC forum phenomenon. We are not able to use high-volume data to do the study analysis. For example, we pick up several representative forum participants and threads to analyze their activity levels and significance, but we do not analyze the whole forum.

Nevertheless, we are among pioneers who are using data analysis to confirm the relation between student forum activity and course scores. As far as we know, we



are the first to put forward definitions and algorithms of super degrees and employ them into MOOC forum analyses. Furthermore, super degrees can reveal more latent or hidden information, which can be applied in social network domains for further research. This study thus has both theoretical and practical applications. In future work, we will try to gather more course data to investigate the relation between forum activities and learning effects, and also employ super degrees parameters into the whole forum to find properties in the context of macro perspective. Furthermore, studying about other parameters in supernetworks is a significant project, as it includes multiple factors such as density, clustering coefficient, and average path length. By employing these parameters into forum study, we may find something more valuable. Finally, social network analysis methods can be used to further investigate participants' behavior in the MOOC discussion forum in future study.

## Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (Grant Numbers: 71373194 and 71573197) and the Fundamental Research Funds for the Central Universities.

## Author Contributions

Chaocheng He (2016201040025@whu.edu.cn) co-wrote the manuscript and analyzed the data. Panhao Ma (841044754@qq.com) collected data, proposed the algorithm and co-wrote the manuscript. Lusha Zhou (437482300@qq.com) searched the literature, analyzed the data and revised the manuscript. Jiang Wu (jiangw@whu.edu.cn) put forward the research problems, designed the research framework and revised the manuscript.

## References

- Alario-Hoyos, C., Perez-Sanagustin, M., Delgado-Kloos, C., Parada, G., H.A., & Munoz-Organero, M. (2014). Delving into participants' profiles and use of social tools in MOOCs. *IEEE Transactions on Learning Technologies*, 7(3), 260–266.
- Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. In the Proceedings of the 23<sup>rd</sup> International Conference on World Wide Web. (pp. 687–698). Seoul, Korea.
- Belanger, Y., & Thornton, J. (2013). Bioelectricity: A Quantitative Approach—Duke University's First MOOC. *Inorganic Materials*, 38(2), 522–526.
- Berge, C., & Miniéka, E. (1973). *Graphs and hypergraphs* (Vol. 7): North-Holland publishing company Amsterdam.
- Breslow, L., Pritchard, D.E., DeBoer, J., Stump, G.S., Ho, A.D., & Seaton, D.T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment*, 8, 13–25.



**Research Paper**

- Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of Interactive Media in Education*, 2012(3).
- Downes, S. (2010). New technology supporting informal learning. *Journal of Emerging Technologies in Web Intelligence*, 2(1), 27–33.
- Edmunds, A., & Morris, A. (2000). The problem of information overload in business organisations: A review of the literature. *International Journal of Information Management*, 20(1), 17–28.
- Engle, D., Mankoff, C., & Carbrey, J. (2015). Coursera's introductory human physiology course: Factors that characterize successful completion of a MOOC. *The International Review of Research in Open and Distance Learning*, 16(2), 46–68.
- Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *Internet & Higher Education*, 23(5), 18–26.
- Ho, A.D., Reich, J., Nesterko, S.O., Seaton, D.T., Mullaney, T., Waldo, J., & Chuang, I. (2014). HarvardX and MITx: The first year of open online courses, Fall 2012–Summer 2013.
- Huang, J., Dasgupta, A., Ghosh, A., Manning, J., & Sanders, M. (2014). Superposter behavior in MOOC forums. In the Proceedings of the 1<sup>st</sup> ACM Conference on Learning@scale (pp. 117–126). Atlanta, Georgia.
- Hull, D.M., & Saxon, T.F. (2009). Negotiation of meaning and co-construction of knowledge: An experimental analysis of asynchronous online instruction. *Computers & Education*, 52(3), 624–639.
- Kellogg, S., Booth, S., & Oliver, K. (2014). A Social network perspective on peer supported learning in MOOCs for educators. *International Review of Research in Open & Distance Learning*, 15(5), 263–289.
- Kizilcec, R.F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In the Proceedings of the 3<sup>rd</sup> International Conference on Learning Analytics & Knowledge (pp. 170–179). Leuven, Belgium
- McAuley, A., Stewart, B., Siemens, G., & Cormier, D. (2010). The MOOC model for digital practice. Retrieved from [https://oerknowledgecloud.org/sites/oerknowledgecloud.org/files/MOOC\\_Final.pdf](https://oerknowledgecloud.org/sites/oerknowledgecloud.org/files/MOOC_Final.pdf)
- Nagurney, A., & Dong, J. (2002). *Supernetworks: Decision-making for the information age*. Northampton, MA: Elgar Edward Publishing, Inc.
- Rossi, L., & Gnawali, O. (2014). Language independent analysis and classification of discussion threads in Coursera MOOC forums. In the Proceedings of the IEEE 15<sup>th</sup> International Conference on Information reuse and Integration. San Francisco, CA.
- Sheffi, Y. (1985). *Urban transportation network. Equilibrium analysis with mathematical programming methods*. Upper Saddle River, NJ: Prentice Hall.
- Siemens, G. (2014). *Connectivism: A learning theory for the digital age*. Retrieved from <http://www.elearnspace.org/Articles/connectivism.htm>
- Stump, G.S., Deboer, J., Whittinghill, J., & Breslow, L. (2013). Development of a framework to classify mooc discussion forum posts: Methodology and challenges. In the Proceedings of NIPS Workshop on Data Driven Education.
- Yang, D., Sinha, T., Adamson, D., & Rose, C.P. (2013). “Turn on, Tune in, Drop out”: Anticipating student dropouts in Massive Open Online Courses. In the Proceedings of NIPS Workshop on Data Driven Education.



- Yang, D., Wen, M., & Rose, C. (2014). Peer influence on attrition in massively open online courses. In the Proceedings of 7<sup>th</sup> International Conference on Educational Data Mining.
- Zhan, Z., Fong, P.S.W., Mei, H., Chang, X., Liang, T., & Ma, Z. (2015). Sustainability education in massive open online courses: A content analysis approach. *Sustainability*, 7(3), 2274–2300.
- Zhao, H., Sullivan, K.P., & Mellenius, I. (2014). Participation, interaction and social presence: An exploratory study of collaboration in online peer review groups. *British Journal of Educational Technology*, 45(5), 807–819.



This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

