Investigate the relationship between learners’ social characteristics and academic achievements

Zhi Liu\textsuperscript{1,2}, Lingyun Kang\textsuperscript{1}, Zhu Su\textsuperscript{1,2*}, Sannyuya Liu \textsuperscript{1,2}, Jianwen Sun\textsuperscript{1,2}

\textsuperscript{1}National Engineering Research Center for E-Learning, Central China Normal University, Wuhan, P. R. China
\textsuperscript{2}National Engineering Laboratory for Technology of Big Data Applications in Education, Central China Normal University, Wuhan, P. R. China

E-mail: suz@mail.ccnu.edu.cn

Abstract. The social interaction behaviors of group learning have obtained a lot of attention from researches, and social network analysis (SNA) plays a critical role in exploring collective interactive patterns. In this paper, we investigated the experimental data from the discussion forum of "Discourse Analysis" course in a university learning platform. By calculating learners social characteristics, we explored the relationship between social characteristics and academic achievements, as well as the differences between various achieving learners. The results show that there exist strong positive correlations between most social characteristics and academic achievements, and the majority of high-achieving learners are located at the core position of the network. In addition, the high-achieving learners exhibited more active participations in discussions and received more posts compared with middle-and low-achieving learners.

1. Introduction

As the focus on online learning continues unabated, Massive open online courses (MOOCs) like Coursera, Udacity, edX and Khan Academy provide learners with better learning support and experience to some extent \cite{1, 2}. To adapt MOOCs to higher education, a variant of MOOCs- Small Private Online Course (SPOCs) have emerged to support the online learning of on-campus students locally \cite{3}. Teachers often conduct interactive teaching through SPOC forums, in which learners can constitute various learning communities for collaborative learning, ideas-sharing, and help-seeking from peers \cite{4}. There is a large amount of data generated in the process of knowledge construction, and social network analysis (SNA) is one of the most important methods in the analysis of data collected through online learning communities. By SNA method, we can understand how the learners interact with each other and what role each learner plays in the network. Besides, by uncovering the interactive factors between learners, learners’ knowledge construction and the development of higher-order thinking may be effectively improved. Therefore, it is worth to investigate that, what the relationship between social characteristics and academic achievements is, and what the social characteristics differences between various achieving learners are. The exploration of these questions is beneficial for teachers timely adjust resource allocation and teaching strategies to provide better personalized services for learners. This paper aims to adopt social network analysis method to carry out
the empirical research to reveal the relationship between social characteristics and academic achievements, as well as the differences of social characteristics between various achieving learners. The remainder of this paper is organized as follows: In Section 2, we review the definition of social network analysis and related research in the online learning environment. Section 3 describes the design of this study. Results are detailed in Section 4, followed by the conclusion in Section 5.

2. Related works
Social network analysis is a method and technique used to study the interactivity among members within the exchange of resources. Members are represented as the nodes in the network, and flow of information between members constitute the links between nodes [5]. SNA has been first applied to the field of mathematics, and then to other fields by sociologists and psychologists in the first half of the 20th century. It provides a quantifiable means for the behavioral characteristics of group learning [6]. A significant number of researches have been dedicated to capture more insight into organizational phenomena through SNA approach. Oleksandra & Shane integrated SNA, discourse content analysis and statistical network modelling to jointly explore the social processes among a group of learners with relatively continuous contributions to a MOOCs course [7]. Wise et al. suggested that different tie definitions and content/non-content networks both have an impact on social network structures in MOOCs discussion forums [8]. Liu et al. adopted SNA to explore how elementary learners collaborate with peers to create stories and discovered that the knowledge level of learners has a certain influence on their position in the network [9]. Laningawijnen et al. utilized SNA to verify whether peer norms for achievement goals have an impact on friendship selections or not [10]. Recently, some researches focused on exploring the relationship between social characteristics and academic achievements. For instance, Williams et al. measured student centrality during the semester to predict future academic performance [11]. Gitinabard et al. adopted learners’ online interaction data establishing the directed network to investigate how the learners’ social characteristics are correlated with academic achievements in three blended courses [12]. Related ideas were pursued by Lee & Bonk. [13] and Liu et al [14]. Houston et al. compared the correlations between course final grade and engagement patterns, and suggested that the number of interactions have stronger correlations with final grade [15]. Joksimovi et al. investigated how the factors generated during learners interactions within two instances on the MOOCs platforms have the influence on the relationship between the three social network centrality measures (i.e., degree, closeness, and betweenness) and final course outcomes [16].

3. Research Design
3.1. Forum Dataset
In this paper, we have collected the interaction data from StarC, which is a learning platform and issued by National Engineering Research Center for E-Learning of Central China Normal University (CCNU). The experimental data comes from the discussion forum within the “Discourse Analysis” course that were delivered on the StarC platform in the Spring of 2016. The course lasts for 4 months and presented by integrating the classroom teaching and online collaborative discussions. In total, there are 134 learners who are registered in this course across a semester, in which 115 learners participated in course discussions and published 1,358 posts. In addition, the academic achievement of each participant is indicated by the final score given at the end of a semester, which is composed of the usual score and the final score. The average score of these learners reaches 83.8 with the scale of 0 ~ 100.
3.2. Research Questions
Investigating the relationship between social interactions and academic achievements contributes to understand the potential causes behind different levels of achievement among learners, thus it is necessary to analyze the interaction patterns between various achieving learners. In this experiment, the discussion posts published within the “Discourse Analysis” course forum will be used to explore the following two aspects: (1) Relationship between social characteristics and academic achievements. (2) Differences on network positions of various achieving learners, as well as the social characteristics differences among various achieving learners. In sum, this paper aiming at addressing the following questions:

1. What is the relationship between social characteristics and academic achievements?
2. What are the differences between network positions of various achieving learners?
3. What are the differences of social characteristics between various achieving learners?

3.3. Instruments and Methods
3.3.1. Social characteristics
The network is composed of individuals and the interactions between them. It is the exchange of resources among individuals that make the information flow from one direction to another. Therefore, it is very necessary to know how individuals in the network control the frequency and direction of communication can enable the correct dissemination of information. SNA provides a quantifiable means, which are conducive to discover what roles individuals play in the network. The quantitative indicators are explained as follows:

| Indicators  | Implications                                                                 |
|------------|-----------------------------------------------------------------------------|
| Authority  | Learners with important influence in the network, it is also can be quantified by indegree centrality of the corresponding node. |
| Betweenness| Learners who on the shortest path connecting two other learners.              |
| Closeness  | The length of shortest path between the learner and others.                  |
| Clustering | It means how well the learners neighbors are connected to each other in the network. |
| Degree     | The frequency a learner communicates with other learners. In this experiment, it is also quantified as the total number of learner posts. |
| Eigenvector| It reveals the importance of learners in the network, and a learner can gain indirect influence from the high-eigenvector learners around him. |
| Hub        | Learners with key links in the network, it is also can be quantified by outdegree centrality of the corresponding node. |
| Indegree   | Number of posts received by a learner.                                      |
| Outdegree  | Number of posts sent by a learner.                                          |
| Pagerank   | A learners weighted popularity, reputation, or influence within a group [11]. |

3.3.2. The division of various achieving learners
Kelley proposed that the optimal choice for dividing upper and lower groups is based on 27% of the score distribution [17], thus we divide all the learners \( (N = 134) \) on the “Discourse Analysis” course into high- \( (N = 36, \text{score} \geq 88) \), middle- \( (N = 62, 81 < \text{score} < 88) \) and low-achieving \( (N = 36, \text{score} \leq 81) \) groups. It is worth to note that not all learners are involved in forum discussions. Among the 115 participants, there are 35 high-achieving learners, 52 middle-achieving learners and 28 low-achieving learners, respectively.

3.3.3. Data Processing
Above all, the data extracted in the “Discourse Analysis” course forum would be filtered and organized, some noisy samples have been removed (e.g., posts made by
teachers in forums, non-registered learners posts). Then, the tool of Python’s NetworkX Package has been used to calculate each participants social characteristics, and the Gephi visual software has been adopted to analyze network positions of various achieving learners. In the end, to demonstrate the relationship between social characteristics and academic achievements, the IBM SPSS software has been used to calculate Spearman correlation and Regression, at the same time, the Kruskal-Wallis H test and Pairwise Comparisons have done to compare the differences of social characteristics between various achieving learners.

4. Research results and analysis

4.1. Descriptive Statistics

It can be seen from Figure 1 that there are 115 learners involved in the discussions and 1,358 posts have been generated, in which 35 high-achieving learners, 52 middle-achieving learners and 28 low-achieving learners generated 527, 561 and 270 posts, respectively. It is worth to note that teachers posts are not included in the statistics. We can also observe that the participation rate of overall forum discussion is 85% and the number of posts per learner is 11:8, in which the different groups participation rate is 97%, 83% and 77%, as well as the number of posts per learner is 14:6, 9:05 and 7:5, respectively. The statistical results indicate that the forum discussion participation rate shows the same downward trend as the number of posts per learner, which further reveals that high-achieving learners participate more actively in forum than others.

4.2. Correlation Analysis and Regression Analysis

To address the first research question, the relationship between learners social characteristics and academic achievements is measured by Spearman Correlation in Figure 2. It shows that most of the social characteristics have significant relationship with academic achievements except for clustering, hub and outdegree. It can be unfolded that the correlation coefficient between the score and authority in this course reaches 0.368 (p < 0.01), indicating that there is a significantly positive correlation between the two variables. The authority can be quantified by indegree centrality of the corresponding node, in other words, the learners with high-authority often receives more posts from others. It makes perfect sense that when a learner has more adequate knowledge about the course content, his/her academic performance would be better,
in fact, high-achieving learners are the most popular ones to be asked for help or discussed problems by others. We also observe that there is a significant positive correlation between score and closeness, which means that the higher the closeness of a learner, the less dependence on others when the learner acquires resources. It also implies that the closer the learner is to others in the network, the better his/her academic achievements will be.

However, it can be seen from Figure 2 that hub and outdegree have weak correlations with academic achievements, which may be attributable to the inherently nature of the course. Actually, there are many differences between foreign language learning and native language learning such as words, the order of sentences, as well as cultural differences. Thus learners need to change their original linguistic thinking in order to further understand the course content, which is obviously more difficult than other non-native language learning courses. Consequently, the teachers of the “Discourse Analysis” program may need to provide targeted help to learners who ask questions in the forum and the learners should also ensure the sufficient time and rich content of the online learning before sending posts online. We acknowledge that these interpretations are speculative, but the context is indeed the primary factor affecting learner behavior patterns [18], as well as the prediction of academic performance with social characteristics.

Then, the linear regression analysis was conducted to examine to what extent the social characteristics are valid indices to predict academic achievements. Among these 10 social characteristics, we only detect that eigenvector has a significant predictive relationship with academic achievements, which is shown in Table 2. Furthermore, we made a multiple regression analysis to verify whether other social characteristics can effectively predict eigenvector centrality or not. For this purpose, using the stepwise selection method to select more relevant factors and eliminate insignificant factors, VIF (Variance Inflation Factor) is used to verify whether there is multicollinearity between the independent variables. It can be seen from Table 3 that the variation range of the VIF value is 1.156 to 4.193 (all less than 10), which means that the multicollinearity is reasonable in this analysis. In addition, in the final linear regression equation, the independent variables explained 97.4% of the dependent variable (corrected $R^2 = 0.974$), $F = 857.532$, $p < 0.001$, which means that the fitting result is very good. The stepwise regression

![Figure 2.](image)
Table 2. Results of linear regression analyses of predicting scores with eigenvector centrality.

| Predictor Variables | B    | SE  | β   | t    | Sig. |
|---------------------|------|-----|-----|------|------|
| (constant)          | 81.698 | .947 | /   | 86.247 | .000 |
| Eigenvector         | 37.376 | 11.488 | .293 | 3.254 | .002 |

Table 3. Results of multiple regression analyses for social characteristics predicting eigenvector.

| Predictor Variables | B    | SE  | β   | t    | VIF |
|---------------------|------|-----|-----|------|-----|
| Degree              | .006 | .000 | .844 | 27.331*** | 4.193 |
| Outdegree           | .004 | .000 | .225 | 10.316*** | 2.085 |
| Clustering          | .007 | .002 | .055 | 3.419**  | 1.156 |
| Betweenness         | -.259 | .069 | -.074 | -3.741*** | 1.699 |
| Closeness           | .017 | .008 | .055 | 2.055*   | 3.153 |

*p < 0.05; **p < 0.01; ***p < 0.001.

reduced the number of independent variables from 9 to 5, including degree, outdegree, clustering and closeness as significant positive predictors of eigenvector, while the betweenness is the significant negative predictors of eigenvector, and the authority, indegree, hub and pagerank were eliminated from the equation.

4.3. The Difference of Network Positions of various achieving learners

Figure 3. (Color online) Sociogram of different achieving learners.

To address the second research question, we draw the sociogram including various achieving learners distributions in social interactions within the course of “Discourse Analysis”. The results are shown in Figure 3, where the orange nodes denote high-achieving learners, purple nodes denote middle-achieving learners, and green nodes denote low-achieving learners. The number on the node is marked in the order of posting. Moreover, the node size represents nodes degree, which is the sum of replying and delivering postings. It can be observed from Figure 3 that the majority of orange nodes are located at the central positions in the network, while
Table 4. Results of Kruskal-Wallis H test for social characteristics of high-, middle- and low-achieving learners.

| Indicators | \(H\)   | \(P\)  |
|------------|---------|--------|
| Authority  | 9.746** | .008   |
| Betweenness| 6.901*  | .032   |
| Closeness  | 8.361*  | .015   |
| Degree     | 10.976**| .004   |
| Eigenvector| 10.086**| .006   |
| Indegree   | 9.385** | .009   |
| Pagerank   | 8.826*  | .012   |

**\(p < 0.01\); *\(p < 0.05\).

Table 5. Results of Pairwise Comparisons of high-, middle- and low-achieving learners.

| Variable    | Group(I)          | Group(J)          | \(H\)   | \(P\)  |
|-------------|-------------------|-------------------|---------|--------|
| Authority   | Low-achieving     | Middle-achieving  | 3.477   | 0.577  |
|             | Low-achieving     | High-achieving    | 18.832**| .005   |
|             | Middle-achieving  | High-achieving    | 15.355**| .008   |
| Betweenness | Low-achieving     | Middle-achieving  | 4.04    | 0.498  |
|             | Low-achieving     | High-achieving    | 15.707* | .015   |
|             | Middle-achieving  | High-achieving    | 11.667* | .036   |
| Closeness   | Low-achieving     | Middle-achieving  | 3.404   | 0.585  |
|             | Low-achieving     | High-achieving    | 17.532**| .009   |
|             | Middle-achieving  | High-achieving    | 14.128* | .015   |
| Degree      | Low-achieving     | Middle-achieving  | 9.055   | 0.24   |
|             | Low-achieving     | High-achieving    | 26.504**| .001   |
|             | Middle-achieving  | High-achieving    | 17.449* | .015   |
| Eigenvector | Low-achieving     | Middle-achieving  | 10.304  | 0.182  |
|             | Low-achieving     | High-achieving    | 25.939**| .002   |
|             | Middle-achieving  | High-achieving    | 15.636* | .03    |
| Indegree    | Low-achieving     | Middle-achieving  | 3.971   | 0.524  |
|             | Low-achieving     | High-achieving    | 18.743**| .005   |
|             | Middle-achieving  | High-achieving    | 14.772* | .011   |
| Pagerank    | Low-achieving     | Middle-achieving  | 3.9     | 0.532  |
|             | Low-achieving     | High-achieving    | 18.204**| .007   |
|             | Middle-achieving  | High-achieving    | 14.304* | .014   |

* *\(p < 0.01\); *\(p < 0.05\).

most of purple nodes and green nodes are located at the edge positions in the network, which signifies that high-achieving learners are more positive in the forum. We also observe that most arrows are pointed to the center, which implies that the higher degree centrality learners often receives more posts than others.

4.4. Social Characteristics Differences between Various Achieving Learners

To address the third research question, the differences between the social characteristics of three achieving groups are examined. From Figure 2, we can observe that seven social characteristics (authority, betweenness, closeness, degree, eigenvector, indegree and pagerank) have significant relationship with academic achievements. Furthermore, the Kruskal-Wallis H test is utilized
to explore the differences between these seven social characteristics of high-, middle- and low-achieving learners. The results are presented in Table 4, which reveals that various achieving learners significantly differ in these seven social characteristics. For further analysis, a post-hoc test after Kruskal-Wallis is adopted to find out the specific differences between the three achieving groups. As shown in Table 5, we can find that, authority, betweenness, closeness, degree, eigenvector, indegree and pagerank are all significantly different between high- and middle-achieving learners, and especially between high- and low-achieving learners. However, middle- and low-achieving learners do not significantly differ in these seven social characteristics. This can also be seen from Figure 3, in which there is no obvious difference in the distribution of middle- and low-achieving learners in the network. These results further suggested that, the position of learners in the network is correlated with their academic performance, as well as high-achieving group has a more active participation in discussions than middle- and low-achieving groups.

5. Conclusion
In this paper, we have utilized the social network analysis to investigate the differences of social characteristics between different achievement levels of learners within a course forum in a university online learning platform. To begin with, we demonstrated that most social characteristics have significantly positive correlations with academic achievements. In addition, we conducted a linear regression analyses and indicated that eigenvector centrality seemed to be a valid indicator to predict academic achievements, that is, the more positively the learners participate in the forum and communicate with the high-centrality peers, the better his/her academic achievement will be. Moreover, the distributions of network positions of various achieving learners show that the majority of high-achieving learners are located at the core positions in the network, while most middle-achieving and low-achieving learners are both located at the edge positions of the network. This reveals that a higher social centrality may be associated to a higher academic performance [19]. In the end, The Kruskal-Wallis H test and Pairwise Comparisons were used to examine the differences of social characteristics between various achieving learners, which reveals that high-achieving learners tend to be more active than middle-achieving and low-achieving learners in forum interactions (indicated by authority, betweenness, closeness, degree, eigenvector, indegree and pagerank). In future work, we will collect more large-scale data that covering more courses to get in-depth insights into the interaction learning patterns between online learners.

6. Acknowledgments
This work was supported by China Postdoctoral Science Foundation (Grant Nos. 30205010003), self-determined research funds of CCNU from the colleges’ basic research and operation of MOE (Grant Nos. KJ02072017-2291), National Natural Science Foundation of China (Grant Nos. 11505071 and 61702207), and National Key Research and Development Program of China (Grant No. 2017YFB1401303).

References
[1] Xu L, Wang F and Yu B 2018 Social Network Analysis of MOOC Learners Knowledge Building (Singapore: Springer) pp 363-377
[2] Chen Y H and Chen P 2015 MOOC study group: Facilitation strategies, influential factors, and student perceived gains Comput. Educ 86 5570
[3] Fox A 2013 From MOOCs to SPOCs. Communications of the ACM 56 38-40
[4] Al-Rahmi W M, Alias N, Othman M S, Marin V I and Tur G 2018 A model of factors affecting learning performance through the use of social media in malaysian higher education Comput. Educ 121 59-72
[5] Haythornthwaite C 1996 Social network analysis: an approach and technique for the study of information exchange Libr. Infor. Sci. Res 18 323-342
[6] Wey T, Blumstein D T, Shen W and Jordan F 2008 Social network analysis of animal behaviour: a promising tool for the study of sociality Anim. Behav 75 333-344
[7] Oleksandra P and Shane D 2016 Untangling MOOC learner networks Int. Conf. on Learning Analytics and Knowledge (Edinburgh) pp 208-212
[8] Wise A F, Cui Y and Jin W Q 2017 Honing in on social learning networks in MOOC forums: examining critical network definition decisions Int. Conf. on Learning Analytics and Knowledge (Vancouver) pp 383-392
[9] Liu C C, Chen Y C and Tai S J D 2017 A social network analysis on elementary student engagement in the networked creation community Comput. Educ 115 114125
[10] Laningawijnen L, Ryan A M, Harakeh Z, Shin H and Vollebergh W A M 2018 The moderating role of popular peers’ achievement goals in 5th- and 6th-graders achievement-related friendships: a social network analysis J. Educ. Psychol 110 289
[11] Williams E A, Zwolak J P, Dou R and Brewe E 2017 Engagement, integration, involvement: supporting academic performance and developing a classroom social network preprint arXiv:1706.04121
[12] Gitinabard N, Xue L, Lynch C F, Heckman S and Barnes T 2017 A social network analysis on blended courses preprint arXiv:1709.10215
[13] Lee J and Bonk C J 2016 Social network analysis of peer relationship and online interactions in a blended class using blogs Internet. High. Educ 28 35-44
[14] Liu Z, Kang L Y, Domanska M, Liu S and Fang C L 2018 Social characteristics of Learners in a Course Forum and Their Relationship to Academic achievements Int. Conf. on Computer Supported Education (Madeira) pp 15-21
[15] Houston I, S L, Brady K, Narasimham G and Fisher D 2017 Pass the Idea Please: The Relationship between Network Position, Direct Engagement, and Course Performance in MOOCs In Proc of the 4th ACM Conf on Learning @ Scale (Cambridge) pp 295-298
[16] Joksimovi S, Manataki A, Gaevi D, Dawson S, Kovanovi V and De Kereki I F 2016 Translating network position into performance: importance of centrality in different network configurations Int. Conf. on Learning Analytics and Knowledge (Edinburgh) pp 314-323.
[17] Kelley T L 1939 The selection of upper and lower groups for the validation of test items. J. Educ. Psychol 30 17-24
[18] Simmel G and Wolf K H 1950 The Sociology of Georg Simmel (Free Press)
[19] Bakerdoyle K 2013 Social network analysis and education: theory, methods and applications (Sage Publications)