Investigating the Differences of Student Interactions between Behavior- and Content-based Networks in Online Discussions

Tianhui Hu[1], Huanyou Chai[2], Sannyuya Liu[1,2], Qian Zhang[2], Guanxian Yi[2], Zhi Liu[2] and Zhu Su[2,*]
[1]*National Engineering Laboratory for Educational Big Data, Central China Normal University, Luoyu Road 152, 430079 Wuhan, China*
[2]*National Engineering Research Center for E-Learning, Central China Normal University, Luoyu Road 152, 430079 Wuhan, China*

Keywords: Online Discussions, SNA, Behavior-based Networks, Content-based Network, Student Interactions.

Abstract: The online asynchronous forum provides a platform for learners to interact with their peers and furtherly improve their critical skills. Understanding the characteristics of student interactions is thus the key to acquiring some useful insights about how learning occurs in online learning environment. Social network analysis (SNA) as a useful tool is often used to analyze student interactions in behavior-based network, in which network tie is defined as the responsive or co-occurrence relation. However, effective student interactions usually rely on the communication of course content as a form of knowledge, not the behavior itself. To this end, this paper began with the word segmentation of every student's posts, then constructed a network with ties defined as the relations between learners who have co-occurrence of course contents words in their posts, and finally examined the differences of group and individual indexes between behavior- and content-based networks. Results showed that there existed significant differences in the structural and statistical properties between these two networks, and the content-based network was more conducive to discovering the actual interactions between learners in online discussions.

1 INTRODUCTION

With the development of online learning practices, more and more new technologies and applications have been incorporated into universities and other institutions of higher education. As auxiliary platform supporting online learning, online asynchronous discussion forums provide a good learning space for learners to communicate with each other and participate together (Kurnaz et al., 2018). Learners interact through discussion dialogues and knowledge sharing activities, thus promoting the internalization of knowledge and the improvement of cognitive skills. Empirical studies have shown that effective interactions are the key to demonstrating the effectiveness of the discussion forums (Tirado et al., 2015). Therefore, understanding the characteristics of learner interactions contributes to understanding how learning occurs and advances in the online learning environments (Dado and Bodemer, 2017)

Meanwhile, there is a greater call for analysis methods that can generate meaningful insights about learner interactions, as more and more learning processes and outcomes data are stored in the online learning platforms (Dado and Bodemer, 2017). Social network analysis (SNA) is therein one such method that gains widespread attention from relevant researchers. For example, SNA was adopted by Shea et al. (2013) to evaluate how the forms of learning presence relate to the network location of students in these interaction spaces. Another study conducted by Liu et al. (2017) investigated how primary school students collaborated with their peers to create multimedia stories through SNA.

However, the majority of the current studies have constructed social networks based on the ties defined as responsive or co-occurrence relations (Fincham et al., 2018; Wise and Cui, 2018), lacking sufficient

[*] Zhu Su is the corresponding author of this paper.

Hu, T., Chai, H., Liu, S., Zhang, Q., Yi, G., Liu, Z. and Su, Z.
Investigating the Differences of Student Interactions between Behavior- and Content-based Networks in Online Discussions.
DOI: 10.5220/0009346602530260
In *Proceedings of the 12th International Conference on Computer Supported Education (CSEDU 2020) - Volume 2*, pages 253-260
ISBN: 978-989-758-417-6

attention to the nature of student interactions, i.e., the communication of course contents embedded in the posts of online discussions. Basically, not every responsive or co-occurrence interaction involves knowledge construction or the development of cognitive skills. Some studies have demonstrated that student interactions are often shallow (Peters and Hewitt, 2010) and disjointed (Thomas, 2002) in online discussions. Conversely, interactions based on the discussions of same course contents can really reveal the process of knowledge construction and the development of critical skills (Hou and Wu, 2011). Therefore, constructing a network with its ties defined as the communication of course content may contribute to better understanding of the interactions and learning processes among learners. To achieve this goal, this paper proposes a content-based (social) network that defines ties as the relations between learners who have co-occurrence of course contents in their posts. Then the differences of group and individual indexes between behavior- and content-based networks are analyzed in order to validate the effectiveness of the latter network.

The rest of this article will be organized as follows: In Section 2, we review the relevant research that applies SNA into educational field (especially online learning). The design of this study is in Section 3, results can be seen in Section 4 and Section 5 presents the conclusions of this paper.

2 RELATED WORKS

SNA, as its name implies, is to analyze the relationships formed by the interactions between nodes in social networks (Freeman, 2011). It consists of two elements, including nodes and ties. A node is a point that is abstracted with no relation with its shape, size, or properties. It can be an individual, a school, a company, a country, etc. A tie is the connection between nodes, that is, the content of the relationship between two nodes, which can be the transfer of materials, the evaluation between individuals, etc (Tichy et al., 1979).

In general, SNA methods mainly include egocentric and global network analyses (Dado and Bodemer, 2017; Jan et al., 2019). The former is used to describe an individual's personal network, focusing on how individual nodes are embedded in the network and affected by the overall network structure. Corresponding measures are to determine the positions of nodes in the network, mainly including degree centrality, betweenness centrality, closeness centrality and eigenvector centrality. In addition, the latter focuses on the overall network structure by describing the patterns of relations in the network. The indexes at global level mainly include network size, density, and some measures of network attributes, i.e., analyses of cohesion, centralization, reciprocity, and tie strength.

SNA is often used to describe the interactions or relationships between individuals and groups in various fields. Recently, plenty of researchers have adopted SNA as a typical method in learning analytics to analyze student interactions in online learning (Ergün and Usluel, 2016; Erlin et al., 2009; Giri et al., 2014; Liu et al., 2017; López et al., 2014). For example, using SNA, Liu et al. (2017) analyzed the learning process in the online creative community involving complex social network activities among students; Ergün and Usluel (2016) used SNA to evaluate the communication structure in an educational online learning environment to understand student participation levels and interactions over time.

Just as some researchers put it, tie definitions play an important role in analyzing the structural and statistical properties in the generated network (Joksimović et al., 2017). According to Fincham et al. (2018), there are usually two distinct categories of tie definitions; One is based on actual communication among students, the other is based on the co-occurrence participation in the same discussion threads. Correspondingly, five kinds of tie definitions are usually adopted in existing literature: 1) Direct reply, i.e., a tie is constructed when there is a responsive relationship between two learners in the same thread, as shown in Figure 1A; 2) Star reply, i.e., all posts within a thread are considered to be tied to the thread starter, as shown in Figure 1B; 3) Total co-occurrence, i.e., it is assumed that all nodes in the same thread are interconnected, as shown in Figure 1C; 4) Limited co-occurrence, i.e., all nodes are connected to all other ones only in their sub-thread and the thread starter, as shown in Figure 1D; 5) Moving window, i.e., all nodes within a moving window of size N are connected to each other.

In addition, some researchers examined how different tie definitions affect the structure and properties of the generated network. For example, Wise et al. (2017) examined how five kinds of tie definitions impact the structure and properties of the induced network, including direct reply, star reply, direct + star reply, limited co-occurrence and total co-occurrence. Although the findings revealed that the properties of the induced networks are unsusceptible to the tie definitions, they were limited to the descriptive properties without examining the statistical ones, such as interrelationships or homogeneous relationships.

To sum up, existing studies have conducted a lot of discussions on leaner interactions by using tie definition at behavioral level. However, forum interactions are basically the processes of knowledge construction in theory, while the social network based on reply or Star relationship cannot completely reflect the knowledge building process among learners. It is because that learners may just be carrying out pure social communication without in-depth communication on course knowledge. Therefore, this study suggests that the tie definition at behavioral level has some limitations, while adopting the tie definition at content level, i.e., the defining tie as the relations between learners who have co-occurrence of course content words of their posts, can better reflect the forum interaction among learners.

3 EMPIRICAL RESEARCH

3.1 Research Questions

To acquire a better understanding of the characteristics of student interactions in online discussions, this paper, starting from the content of student interactions, constructs a new network using the ties defined as the word co-occurrence after word segmentation of learners' posts. It aims to address two research questions:

(1) What are the differences in group indexes between behavior- and content-based networks?

(2) What are the differences in individual indexes between behavior- and content-based networks?

3.2 Research Objects and Dataset

The data in this study is from a course forum on SPOC platform in a normal university of China. The name of the course is "Freshman Seminar", aiming to help each student of the class in 2018 better integrate into college life and guide them to make appropriate study and career plans. The course is taught by teachers in face-to-face class and additional resources are uploaded to the online platform for students to download and study. In addition, there is a special forum platform for students to communicate and interact online. The course lasts one semester, and there are 133 freshmen, 7 teachers participating in the forum, and 24 senior students (2 seniors, 10 juniors, and 12 sophomores). Finally, 9,798 pieces of data are collected from the forum platform.

By cleaning and screening the forum discussion data, i.e., removing the data of repeated posts, false posts and posts without replies, 8,824 pieces of valid data were finally obtained. Then word co-occurrence network was constructed based on posting contents before calculating the index characteristics (degree centrality index, graph density, etc.) and performing visualization analysis of the network by Gephi 0.9.2. In addition, the behaviour-based network was meanwhile built in order to analyze its differences with the word co-occurrence network constructed based on course content.

3.3 Research Method

In this paper, the python programming language for data processing was adopted using a word segmentation tool called Jieba (a kit in Chinese natural language processing with the affordances of word segmentation, part-of-speech tagging, and named entity recognition) to segment the text of each post with custom segmentation dictionary. After eliminating the corresponding stop words, the content of each post is composed of several words. Then we save the results after word segmentation of each post into a list that is not repeated (if the same word is repeated, it will only be recorded once). In the same thread, compare the previous sender's text data with a certain step size (the step size set in this paper is 10), and set certain conditions to establish a connected relationship. Among them, this paper believes that the connection between two learners should meet the following conditions: the number of co-occurrence words after word segmentation of two contributors exceeds a certain threshold (the threshold of this study is 5) or the number of co-occurrence words after word segmentation of two learners, and the intersection/union set of list is greater than 0.5. In this paper, 426 pieces of data were selected for testing, and the word co-occurrence network extracted based on word segmentation results was compared with the results manually encoded by two researchers. It was found that in the data set of this study, the results with a step size of 10 and a threshold value of 5 were the most consistent with the results manually encoded, reaching 0.74.

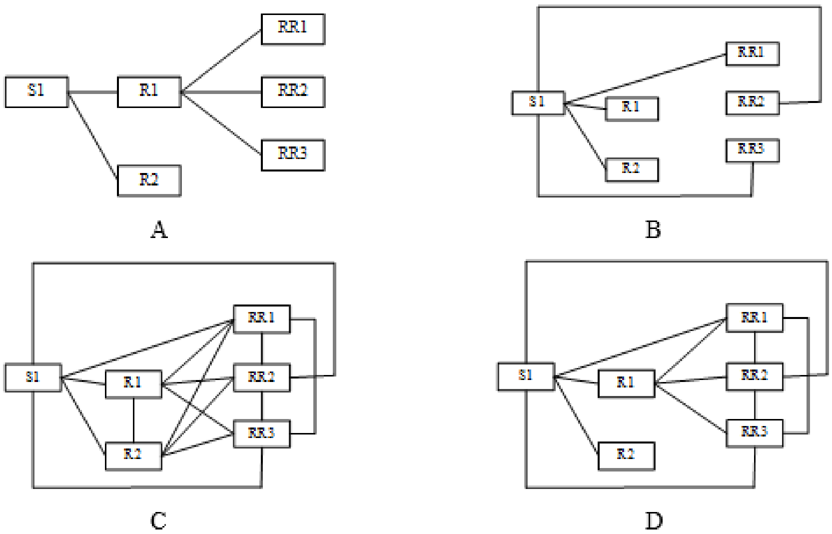

Figure 1: (A) Direct reply;(B) Star reply;(C) Total co-occurrence;(D) Limited co-occurrence.

For the behaviour-based network in the forum, this paper extracted ties according to the star network, and the relationship between learners represents learners' replies or comments to others. Finally, the indicators of two networks are compared to analyze the differences so as to dig deeper into characteristics of learners' interaction patterns.

4 RESEARCH RESULT

4.1 Differences in the Group Indexes between Two Networks

The group index results of behavior- and content-based network are shown in table 1 below. It can be seen that with the exception of the average clustering coefficient, all the indexes of the content-based network are larger than those of the behavior-based network.

Among them, the modularization index of the content-based network is 0.205, which is significantly higher than that of behavior-based network (0.023), indicating that the content-based network is more conducive to discovering the existence of community in the learner groups. The Wilcoxon symbol rank test is then used to compare and analyze the two network indexes, and the result of significance test is P = 0.08 (marginally significant). This indicates that there are some differences between the behavior- and content-based networks, which reflect that traditionally behavior-based network can reflect the interaction at behavioral level, but cannot reflect the implicit connection established by learners in knowledge processing or cognitive thinking.

Figure 2(A) and (B) are the network diagrams for visualizing the two networks. Among them, there are 8656 ties in the behavior-based network with 164 participants, including 7 teachers and 24 senior students. And 7249 network ties are constructed in the content-based network, involving 158 learners. It can be seen that the numbers of ties and nodes in the content-based network are smaller than those in the behavior-based network. Among the 6 participants missing in the content-based network, 3 are course teachers and 3 are freshmen. When they post in the forum, they only participate in the interaction, but do not talk about knowledge or communicate cognitively. Their speech content is relatively simple, such as "thank you", "not quite understand", "I think it is ok". The node size in the figure represents the value of degree centrality of learner, the red node represents the teacher, the green one represents the freshmen, the yellow dot represents the seniors, the purple represents the juniors, and the blue represents the sophomores. In the behavior-based network, teachers are at the central position, and their degrees are relatively large, indicating that teachers often act as the initiators of topics in the forum interaction to promote the communication and discussion among students. In the content-based network, teachers are in a relatively marginal position. It can be seen that although teachers organize the communication of students, they do not play a strong leadership role in knowledge sharing and cognitive improvement.

Table 1: Group indicators of the two networks.

	Degree	Network diameter	Density	Modularity	Average clustering coefficient
Behavior-based network	20.835	3	0.128	0.023	0.324
Content-based network	29.892	4	0.19	0.205	0.304

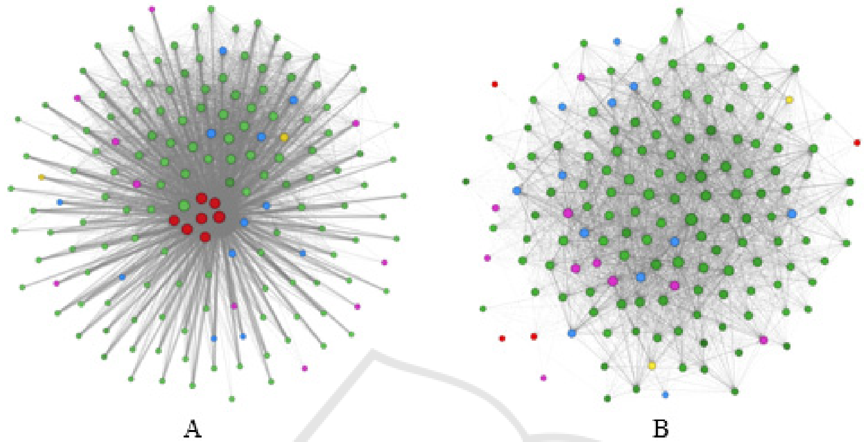

Figure 2: (A) Behavior-based network, (B) Content-based network. The size of nodes represents the value of degree centrality of individual nodes, red nodes represent teachers, yellow nodes represent the seniors, purple nodes represent the juniors, blue nodes represent the sophomores, green nodes represent the freshmen.

Table 2: Descriptive and T-test analysis of individual measures in two networks.

	Behavior-based network (Mean ± SD)	Content-based network (Mean ± SD)	T value
In-degree centrality	18.79 ± 28.88	29.89 ± 13.69	-4.03
Out-degree centrality	21.55 ± 18.97	29.89 ± 13.76	-5.40
Closeness centrality	0.54 ± 0.16	0.53 ± 0.07	0.85
Betweenness centrality	73.03 ± 214.88	134.48 ± 121.23	-3.05
Eigenvector centrality	0.17 ± 0.23	0.46 ± 0.21	-10.71

In addition, from the network structure, we can find that the two networks are quite different. The structure of the behavior-based network looks more like a star in its appearance, largely because of the limited organization of the forum in which one person posts and the others make a reply. In the content-based network constructed on the course content, it can be seen that students interact more with their peers in terms of ideas or cognition. Therefore, we believe that constructing a content-based network from the course content perspective can better reveal learners' interaction in the process of sharing ideas and knowledge, which are hidden in the behavior-based network.

4.2 Differences in the Individual Indexes between Two Networks

Eliminating 6 learners who don't appear in the content-based network, the remaining 158 learners in two networks are analyzed using paired sample T test to examine their differences in in-degree, out-degree, closeness, betweenness, and eigenvector centralities. The results revealed that there are significant differences in all the centrality measures except closeness centrality, as shown in Table 2.

To get a deep understanding of the differences of individual measures in these two networks, two students (S225 and S306) are adopted as an example to illustrate the individual index differences, as shown in table 3. Specifi- cally, all the five indexes of learner S225 in the behavior-based network are higher than average, while in the content-based network, all indexes of this learner are lower than the average except for the closeness centrality. For learner S306, all the indexes in the behavior-based network are lower than the average, while in the content- based network, all the indexes are higher than the average except for eigenvector centrality.

In order to more intuitively show the differences of interaction pattern of learners in the two networks, this study extracted the ties that connect these two learners and other ones for further analysis and visual presentation. From Figure (A) and (B), it can be intuitively seen that, although S225 has a high degree of activity and influence for interaction in the behavior network, in the content-based network, the number of interactions with other learners is significantly less, indicating that its influence on peers is not obvious. Conversely, from Figure 3 (C) and (D), S306's interaction with other learners in the behavior-based network is not as active as that of S225, but it is highly motivated and has a high reputation in this content-based network.

Table 3: Comparison of individual indicators in behavior-based and content-based network.

	Types of network	In-degree	Out-degree	Closeness centrality	Betweenness centrality	Eigenvector centrality
S225	Behavior-based network	71	41	0.68	354.07	0.54
S225	Content-based network	13	14	0.50	17.85	0.22
S306	Behavior-based network	0	16	0.52	0	0
S306	Content-based network	34	37	0.56	158.25	0.48

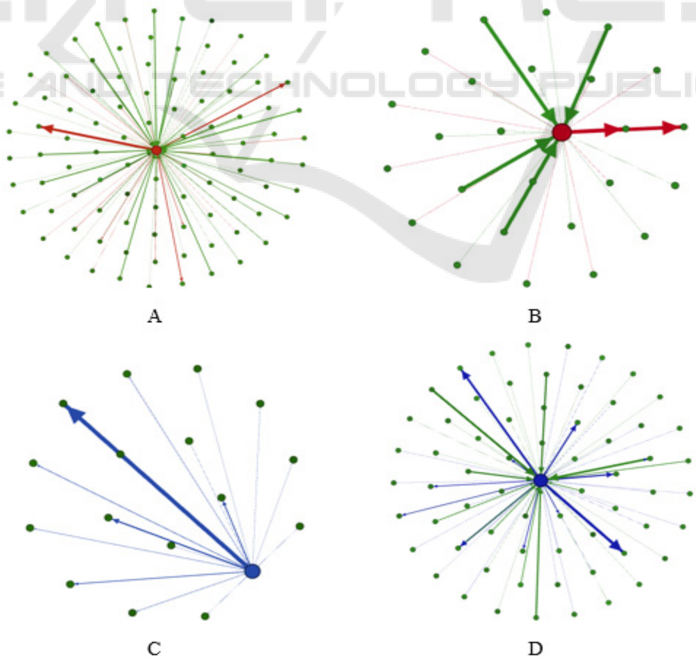

Figure3: (A) and (B) respectively show the interaction patterns of learner S225 in the behavior- and content-based networks. The red nodes represent student S225. Figure (C) and figure (D) respectively show learner S306's interaction patterns with other learners in the behavior- and content-based networks.

5 CONCLUSIONS

This study utilized the method of SNA to examine the interaction patterns of students in online discussions, aiming to explore the actual interactions between learners and further provide some useful insights of effective online education. For most of the current research that define ties as responsive relations at behavior level, neglecting the actual interactions based on course content or knowledge, this paper defines ties as the relations between learners who have co-occurrence of course contents in their discussion posts and further construct a content-based network. Firstly, a tool package for word segmentation called *Jieba* was adopted to segment each learner’s posts extracted from the online discussions. Second, if the word intersection ratio between two posts from two distinct learners is greater than 0.5 or the number of co-occurrence words is greater than 5, a tie would be considered to exist in these learners. Third, based on the above ties, a new network was constructed, different from traditionally behavior-based network. Finally, the differences in group and individual indexes were compared between behavior- and content-based networks.

Compared to the behavior-based network, the number of ties in content-based network is relatively small, but other indexes, including density, modularity and degree in the latter network are higher than those in the former one. These results indicate that learners are more cohesive in the content-based network. While in the behavior-based network, the average clustering coefficient and the average path length index are relatively high, indicating that the distances between learners are relatively far and they tend to establish connections with some influential nodes. In this study, due to the curriculum and the structure of the course and forum factor, most students simply reply to the teacher or assistant and the thread starter. Although behavior-based network to a certain extent can reflect the interactions between the student groups (García-Saiz et al., 2013), it cannot comprehensively show actual interaction relationship between learners. Conversely, the content-based network, defining ties based on the co-occurrence of course content or knowledge, can better reveal the interaction patterns between learners at the cognitive level. In addition, comparing the individual indexes in the two networks, this paper found that the member distributions in the behavior- network and content-based networks changed greatly. These results indicated that there were a large number of shallow level interactions in the forum interaction, that is, learners posted a lot on the platform but lack of knowledge and cognitive interaction with other learners, and they simply replied to the posts under the existing forum structure. Therefore, content-based network could better reflect the implicitly real interactions between learners.

Based on our findings, we could get some useful insights about how to adopt SNA to analyze student interactions in an appropriate manner. As the traditionally behavior-based network cannot fully reveal the actual interactions of learners, the content-based network can to some extent to make up for this defect. First,teachers can use the information of the content-based network to dig out the actual interaction pattern of students in online discussions.Second, teachers could encourage students to communicate about the content of knowledge.and create meaningful viewpoints to promote students' knowledge construction when guiding students.

ACKNOWLEDGEMENT

This work was supported by Program of National Natural Science Funds of China [Grant No. 61977030], the Fundamental Research Funds of the Central Universities [Grant No. 2019YBZZ007, CCNU19ZN012] and the China Postdoctoral Science Foundation (Grant No. 3020501003). No competing financial interests existed.

REFERENCES

Dado, M., Bodemer, D. (2017). A review of methodological applications of social network analysis in computer-supported collaborative learning. *Educational Research Review,* 22: 159–180.

Ergün, E., Usluel, Y. K. A. (2016). An analysis of density and degree-centrality according to the social networking structure formed in an online learning environment. *Educational Technology & Society,* 19(4): 34–46.

Fincham, E., Gašević, D., Pardo, A. (2018). From social ties to network processes: do tie definitions matter? *Journal of Learning Analytics,* 5(2): 9–28.

Freeman, L. C. (2011). The development of social network analysis–with an emphasis on recent events. *The SAGE handbook of social network analysis*, 21(3): 26–39.

García-Saiz, D., Palazuelos, C., Zorrilla, M. (2013). Data mining and social network analysis in the educational field: an application for non-expert users. *In*

Educational data mining: Applications and trends (pp. 411–439). Berlin, Heidelberg: Springer, Berlin/Heidelberg.

Giri, B. E., Manongga, D., Iriani, A. (2014). Using Social Networking Analysis (SNA) to analyze collaboration between students (Case study: Students of open university in kupang. *International Journal of Computer Applications (0975-8887),* 85(1): 44–49.

Hou, H., Wu, S. (2011). Analyzing the social knowledge construction behavioral patterns of an online synchronous collaborative discussion instructional activity using an instant messaging tool: A case study. *Computers & Education,* 57(2): 1459–1468.

Hyo-Jeong, S. O. (2014). Towards rigor of online interaction research: Implication for future distance learning research. *Distance Education in China,* 9(2): 256–263.

Jan, S. K., Vlachopoulos, P., & Parsell, M. (2019). Social network analysis and learning communities in higher education online learning: A systematic literature review. *Online Learning Journal*, 23(1): 249–264.

Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., et al. (2017). How do we model learning at scale? A systematic review of research on MOOCs. *Review of Educational Research,* 88(1): 43–86.

Kurnaz, F. B., Ergün, E., Ilgaz, H. (2018). Participation in online discussion environments: Is it really effective? *Education & Information Technologies,* (3): 1–18.

Liu, C., Chen, Y., Diana Tai, S. (2017). A social network analysis on elementary student engagement in the networked creation community. *Computers & Education,* 115: 114–125.

López, P. M., Aliaño, A. M., Gómez, J. I. A. (2014). Social Network Analysis of a Blended Learning experience in Higher Education. *Research on Education and Media,* 6(2): 69–78.

Peters, V. L., Hewitt, J. (2010). An investigation of student practices in asynchronous computer conferencing courses. *Computers & Education,* 54(4): 951–961

Shea, P., Hayes, S., Smith, S. U., Vickers, J., Bidjerano, T., Gozza-Cohen, M., & Tseng, C. H. (2013). Online learner self-regulation: Learning presence viewed through quantitative content-and social network analysis. *The International Review of Research in Open and Distributed Learning*, 14(3): 427–461.

Thomas, M. J. W. (2002). Learning within incoherent structures: the space of online discussion forums. *Journal of Computer Assisted Learning,* 18(3): 351–366.

Tichy, N. M., Tushman, M. L., & Fombrun, C. (1979). Social network analysis for organizations. *Academy of management review,* 4(4): 507–519.

Tirado, R., Hernando, Á., & Aguaded, J. I. (2015). The effect of centralization and cohesion on the social construction of knowledge in discussion forums. *Interactive Learning Environments*, 23(3): 293–316.

Wise, A. F., Cui, Y. (2018). Learning communities in the crowd: Characteristics of content related interactions and social relationships in MOOC discussion forums. *Computers & Education,* 122: 221–242.

Wise, A. F., Cui, Y., & Jin, W. Q. (2017, March). Honing in on social learning networks in MOOC forums: Examining critical network definition decisions. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp.383–392). ACM.

Yusof, N., & Rahman, A. A. (2009, April). Students' interactions in online asynchronous discussion forum: A Social Network Analysis. *In 2009 International Conference on Education Technology and Computer* (pp. 25–29). IEEE.