Unfolding knowledge co-construction processes through social annotation and online collaborative writing with text mining techniques

Sandy C. Li, Tony K. H. Lai
Department of Education Studies, Faculty of Social Sciences, Hong Kong Baptist University

Despite the positive claims on the pedagogical use of social annotation and online collaborative writing tools discussed in the literature, most of the findings are derived from interviews or self-reported survey data. Very few studies probed deep into the learning processes and examined students' digital traces and the artefacts they co-construct. In this study, we employed semantic network analysis techniques to examine how the use of a social annotation tool (Diigo) coupled with an online collaborative writing (Google Docs) affects students' learning outcomes. The results indicate that the use of Diigo coupled with Google Docs helps enhance student engagement in the collaborative process and that the concept connectivity and quality of the text co-constructed by each group using Diigo coupled with Google Docs is significantly higher than those using Moodle's forum. In addition, the level of collaboration within a group correlates positively with the number of vertices with high lexical relevancy identified in the semantic network of the text co-constructed by each group.

Implications for practice and policy:
- Undergraduate students can use Diigo coupled with Google Docs to enhance their collaborative work.
- Course leaders could use Diigo coupled with Google Docs to support learning activities, such as flipped learning or collaborative inquiry learning, in which students are required to engage in close reading and the co-construction of artefacts.
- Course instructors could consider using semantic measures such as the number of clusters and betweenness centrality to assess the quality of students' co-constructed artefacts.

Keywords: social annotation, collaborative writing, knowledge co-construction, semantic network analysis

Introduction

Social annotation for learning

Bookmarking and annotating text have long been regarded as effective strategies for facilitating close reading and deep learning (Miettinen et al., 2003; Nokelainen et al., 2005). These learning strategies enable readers to encode, decode and transmute useful information during reading. Apart from providing aids to memory, annotations help readers advance their understanding to a higher level by engaging them in a more in-depth reflection on the selected information. With the advancement in social media, a number of popular social annotation technologies have been developed over the last 2 decades. These include EDUCOSM (Miettinen et al., 2003; Nokelainen et al., 2005), Diigo (Gao, 2013; Li et al., 2015), HyLighter (Mendenhall & Johnson, 2010) and Hypothes.is (Dean, 2015). With the emergence of these online social annotation platforms, the physical barriers and cognitive loading imposed by paper-based annotation when users want to collate information or retrieve notes from different types of media can be alleviated (Li et al., 2015).

These online annotation systems help to create a sense of community in which users can share their work and build on that of others (Kear et al., 2016). More importantly, as Johnson et al. (2010) argued, the provision of a highly granular threaded discussion mechanism anchored to specific sections or paragraphs or even sentences of a text in social annotation systems has positive impacts on student engagement in the collaborative process. Nokelainen et al. (2005) conducted a study on using EDUCOSM to support students in learning statistics. The results derived from their pre-test and post-test and student activity logs indicate that (a) students' motivation and social ability predicted their activity in the system, (b) students' activity related positively to students' performance and final grades and (c) all participating students claimed that

the use of the system had a positive impact on their study habit and learning process. Nonetheless, Nokelainen et al. (2005) did not include a control group in their design. This kind of simple repeated-measure design may pose threats such as maturation threat and history threat to its internal validity. Thus, it would be hard to conclude that the observed learning outcomes were due to the intervention. Furthermore, the pre- and post-tests adopted in their work were two independent surveys measuring different constructs. The pre-test measured students' self-reported motivation, learning strategies and social ability while the post-test measured students' perception of the intervention. In other words, the test scores basically cannot be used to measure the gain in student learning outcomes resulting from the intervention.

Likewise, Johnson et al. (2010) conducted two studies on the use of HyLighter to enhance student engagement in reading selected essays. Students were required to complete reading comprehension tests on 5 given articles under different experimental conditions. In terms of students' comprehension skills and metacognition, their pre-test and post-test results indicate that the groups which were assigned to work in a collaborative setting outperformed their counterparts who worked on an individual basis. Their test results also indicate that social annotation helps enhance students' critical thinking skills (Mendenhall & Johnson, 2010). Nonetheless, Mendenhall and Johnson indicated that the major difference between the individual treatments and the group treatments was the ability of the groups who were sitting physically in the same location during the activities to talk about and share ideas. They did not make direct observation of students' performance on the highlighting and annotating activities. As such, it would be hard to conclude if the outperformance of those receiving group treatments was due to their collaborative activities through social annotation or students' face-to-face verbal explanations and peer talks. They pointed out that one of the confounding factors of their study was the limited exposure time (the time for reading an article) to the intervention, which could have impinged on the instructional effect. They suggested that future study designs should enable researchers to observe students' social annotation activities over prolonged periods of time, as well as devise appropriate measures on student learning processes and group collaboration and examine how these measures impede student learning outcomes. Gao (2013) found that students exhibited a wide range of behaviours, including self-reflection, elaboration, internationalisation and the provision of social support when they were learning an online text with Diigo. Gao then conducted a survey to gauge students' perceived learning experience. Although students generally held a positive attitude towards the use of Diigo for collaboration, some of them found it cognitively demanding in navigating through a huge volume of annotations and notes while reading the text and inconvenient to use as it and lacked appropriate facilities for them to organise their notes and to engage in extensive collaborative writing based on the annotated texts and notes.

Technology-supported collaborative writing

Online collaborative writing tools provide a rich platform for facilitating peer scaffolding and knowledge co-construction, which are pivotal to enhancing students' motivation and engagement in the collaborative process. Among various online collaborative writing tools, wiki has been one of the widely studied social media. Nonetheless, the research findings related to the pedagogical use of wikis are equivocal. Neumann and Hood (2009) conducted a quasi-experimental study on using a wiki to support collaborative learning in higher education. Their findings suggest that the wiki approach enhances student engagement, but it makes no impact on student performance. Likewise, Chu et al. (2017) examined the use of wiki to facilitate project-based learning for students from different disciplines at tertiary level. Their findings reveal that students were generally positive towards the use of wiki for their learning, but there were significant differences in students' perceived overall learning experience, motivation, group interaction, engagement and ease of use of the wiki. They also found that wiki's affordance for knowledge management vary significantly among students of different disciplines. In comparison to wikis, Olenewa et al. (2017) pointed out that the ability to write simultaneously in a shared document and to visualise the evolution of the document development process in some online collaborative writing tools like Google Docs has a strong impact on students' learning motivation. The high visibility of collaborators' activities in Google Docs makes participants feel compelled to see and emulate what their group members are doing and ensures that everyone is on the same page and able to work towards the common goals. S. H.-J. Liu and Lan (2016) conducted a study to examine the differences in learning motivation and vocabulary gain between two groups of English-as-a-Foreign Language (students who were randomly assigned to work on a task either individually or in a collaborative setting using Google Docs. The results indicate that, in terms of vocabulary gain, those who worked collaboratively outperformed those who worked individually and that the former exhibited higher learning motivation and better perceived learning experiences than the latter.

In another study, Zhou et al. (2012) found that the use of Google Docs in an out-of-class collaborative writing activity changed the means of communication used by the participants during their group work. Over 90% of the participants regarded Google Docs as a useful tool for collaboration though it had no effect on their final grades. Suwantarathip and Wichadee (2014) and Seyyedrezaie et al. (2016) reported that students held positive attitudes towards the use of Google Docs to support collaborative work and that they generally exhibited a higher level of collaboration in out-of-class collaborative writing with it in comparison to in-class face-to-face collaborative writing activity without it. Along a similar vein, M. Liu et al. (2018) used an online collaborative writing tool, Cooperpad, which has a built-in component for visualising participants' engagement intensity, to facilitate collaborative writing tasks. Their results indicate that both the quality of writing and student engagement of the group using Cooperpad were higher than those of the group using an online collaborative writing tool without a visualisation component. However, the data associated with student engagement was based on students' responses to a post-intervention feedback survey. The relationship between student learning outcomes and the activities in which they engaged remains unknown.

Research gaps

With regard to the studies mentioned above, there are a number of design and methodological issues concerning social annotation research remaining to be addressed. For instance, there is a need to augment the social annotation environment with collaborative writing facilities that enable students to better organise their notes and extend their thoughts beyond their annotations. In addition, to better understand if the use of an augmented social annotation environment can improve student learning, it is appropriate to adopt research designs such as quasi-experimental design or cross-over repeated measure design to enable researchers to compare the effects of the intervention against baseline measurements or against the effects of other pedagogical approaches. In addition, the majority of the above studies rely on perception or self-reported surveys to gauge students' feedback on the use of technology in learning and very few of them probed directly into students' learning processes (Chu et al., 2017). To deepen our understanding of how students' learning trajectories progress, we deemed that a more detailed examination of the quality of the artefacts co-constructed by students during collaboration was necessary in our study.

To address the above issues, firstly, we extended the learning environment by coupling social annotation with an online collaborative writing platform and attempted to examine the entire student learning process from collaborative close reading to co-construction of artefacts over a prolonged period of time. Secondly, we adopted a crossover repeated measure design in our study to enable us to compare student learning processes and learning outcomes in the two different settings. Participants were randomly assigned to two independent groups and each group was required to immerse in two settings – Diigo coupled with Google Doc and Moodle's forum – to facilitate their collaborative work. Thirdly, we employed a quantitative measure to gauge the level of collaboration within a group and used it to examine how group processes affected student learning outcomes. Fourthly, we adopted a semantic network analysis approach to gauge the quality of students' constructed artefacts. Although grades are commonly used as a proxy for assessing students' artefacts, they involve, to a certain extent, instructors' subjective judgement in making qualitative assessment of the artefacts. Alternatively, Hoser et al. (2006) and Haya et al. (2015) employed text mining techniques such as semantic network analysis to assess the quality of students' artefacts based on various semantic measures. They found these measures to be a promising set of tools for providing deep insights into the structure and quality of ontologies or concept words constructed by learners. Further, based on these measures, researchers can derive useful learning analytics that help inform the design of automated adaptive support and scaffolds to learners.

Context of the study

As discussed above, despite the positive claims on the pedagogical use of social annotation and online collaborative writing tools discussed in the literature, most of the findings are derived from interviews or self-reported survey data. Very few studies probed deep into the learning processes and examined students' digital traces and the artefacts they co-constructed. In this study, we attempted to investigate how the use of social annotation coupled with online collaborative writing affects students' learning trajectories, that is, the progression of student learning manifested by the postings, highlighted texts, annotations, comments and content words contributed by each individual at different junctures of the collaboration process. The research questions of this study were twofold:

- In comparison to online discussion forum, how does the use of social annotation tools coupled with online collaborative writing affect students' engagement and the quality of their co-constructed artefacts? Specifically, how does students' engagement as reflected by the number of content words they contributed and the semantic richness of their con-constructed artefacts using social annotation coupled with online collaborative writing compare with that using online discussion forum?
- How does the level of group collaboration correlate with the quality of students' work as reflected by the semantic measures identified in their co-constructed artefacts?

In the following section, we will delineate our research design and the approaches to data analysis.

Methods

To address the above research questions, we adopted a crossover repeated measure design and asked the participants to use different technologies to facilitate their collaborative work. Specifically, we examined how the use of Diigo coupled with Google Docs in comparison to the use of Moodle's forum impacted on group collaboration. Further, by making use of semantic network analysis techniques, we scrutinised how the level of collaboration within a group affected the quality of participants' work. We conducted the study during the second semester of the 2015–2016 academic year, after obtaining ethical clearance from our university.

Participants

A purposive sample of 27 (20 females and 7 males) undergraduate students aged between 19 and 20 from a local university in Hong Kong were recruited in this study. The participants were second-year double-degree students majoring in both English Language as well as Literature and English Language Teaching. They were all enrolled in a major core course on educational technology, which aims to expose students to a variety of emerging educational technologies and how they can be integrated into daily classroom practices. For the sake of convenience, we will use the terms *participants* and *students* interchangeably in this and the following sections. We anticipated that, through their first-hand experiences, students could develop a better understanding of the affordance of different technologies and the major issues concerning classroom integration from both the teacher and student perspectives. All the participants were Chinese with Cantonese as their mother tongue and English as their second language. In this study, students were required to engage in collaborative story analysis with the support of learning technologies. As such, the participants were expected to possess a certain level of English and technology literacies. Thus, the rationale for the above choice of participants were twofold: (a) English literature is an integral part of their formal curriculum, which implies that the participants have a good understanding of story critique and analysis and (b) as they were all enrolled in a major core course in educational technology, they were familiar with the use of different technologies, such as online forums, collaborative concept mapping tools, social annotation tools, blogs and wikis to support collaborative learning. All students were informed of the purpose of the study, and their prior consent for participation was obtained. They were free to opt in or out of the study without causing any prejudice to their course assessment scores.

Research design

The participants were randomly assigned to groups of 4 or 5 (see Table 1) using Moodle's built-in function for random grouping. Although random assignment has the advantage of maximising the heterogeneity of the group, the choice of group size influences the group dynamics and the quality of the group work (Burke, 2011). In this study, the choice of a group size of 4 or 5 was to ensure a sufficient number of individuals to generate multiple perspectives and ideas.

Table 1
Group distribution and the tasks assigned to each set

Group	Set A			Set B		
	A1	A2	A3	B1	B1	B1
Group size	5	4	4	5	5	4
Story 1	Using Moodle's forum as a platform for story analysis			Using Diigo coupled with Google Docs as a platform for story analysis		
Story 2	Using Diigo coupled with Google Docs as a platform for story analysis			Using Moodle's forum as a platform for story analysis		

The groups were further separated into two sets randomly: Set A and Set B. Each group was required to conduct two trials of a story analysis on two given short stories – Story 1 and Story 2 – with the support of different collaborative technologies. The reasons for adopting the crossover repeated measure design are both pedagogical and methodological. Firstly, the story analysis activity was conducted in a face-to-face classroom setting, we wanted to ensure that all the groups from either Set A or Set B were given an equal opportunity to engage in collaborative learning with the two given types of online settings – (a) online discussion forum and (b) social annotation coupled with online collaborative writing – and that their performance would be assessed on an equal footing. Secondly, this crossover repeated measure design can help avoid the situation where the control group and experimental group have some fundamental difference that may distort the results. This situation may arise when the sample groups are small despite random assignment of participants. As shown in Table 1, for Story 1 Trial, the groups associated with Set A were assigned Moodle's built-in discussion forum to facilitate their collaborative story analysis while those associated with Set B were assigned Diigo coupled with Google Docs to support their collaborative work. For the Story 2 Trial, the choices of technologies for Set A and Set B were interchanged. The story analysis task for each trial spanned about 2 weeks. During that time, students were free to choose how much time they wanted to devote to the task.

The two short stories were chosen so as to ensure comparable length and level of difficulty, verified by the one-way between-group ANOVA of the difference in students' performance between Set A and Set B; the details are elaborated in the Results section. In the story analysis task, each group was required to analyse and discuss the basic elements of the story: (a) setting, (b) conflict, (c) character, (d) theme, (e) plot, (f) climax, (g) dialogue and (h) lesson learned. Based upon their discussion, each group was required to compose collaboratively a summary report of their findings.

Analysis of the text contributed by individual members

To delineate the knowledge co-construction process and to compare the collaborative work delivered via Moodle's forum with that delivered via Diigo coupled with Google Docs, each participant's artefacts constructed separately in two different online settings were collected for analysis. For the story analysis task with Moodle's forum, each group member's contribution to the co-construction process was examined based on the notes they posted. For the story analysis task with Diigo coupled with Google Docs, both the notes posted in Diigo and the words contributed by each group member in the shared Google Docs file were filtered out for analysis. To clean up the text contributed by each individual member, Feinerer and Hornik's (2020) stopword list for natural language processing was adopted. After filtering out all the stopwords, the inputs of individual members in the two given online settings (Moodle forum vs Diigo coupled with Google Docs) were compared by counting the number of content words they contributed. This number was used as a measure for gauging their engagement in learning.

Semantic network analysis of students' co-constructed text

Semantic network analysis is a text mining technique derived from social network analysis. In the graphic form of a semantic network, vertices represent concepts found in a specific text corpus while an edge connecting a pair of vertices represents their co-occurrence in close proximity to each other. The number of edges connecting two concepts represents the frequency of their co-occurrence. Researchers could create a window with a specific size to extract word pairs to form a co-reference list in which each pair of words co-occur with one another within a specified window size. The window size can be chosen as one sentence, two sentences, a paragraph or even a few words. With this co-reference list, researchers can construct a proximity matrix. Each entry in the matrix represents the number of edges connecting a pair of words in

the text corpus. The proximity matrix contains all the information for constructing the semantic network. As mentioned earlier, the semantic measures derived from the network provide an objective way to assess the complexity, concept connectivity and lexical relevancy of a given text corpus. These measures are useful for discerning the structure and quality of concepts constructed by students (Haya et al., 2015; Hoser et al., 2006).

To compare the quality of the collaborative work delivered, we constructed semantic networks for the summary reports co-constructed by the groups in the two given online settings and used them to examine the semantic richness of students' work (Haya et al., 2015). To this end, we conducted a semantic network analysis of their reports with a text mining package, AutoMap (Carley et al. 2013). We imported two sample text corpuses comprising all the group story analysis reports associated with the two given short stories separately to AutoMap for analysis. The overall process of semantic analysis basically involved the following phases: (a) text pre-processing, (b) text processing and (c) data analysis.

Pre-processing phase

In this phase, the two text corpuses were cleaned and standardised through a series of steps. These consist of (1) de-capitalisation, (2) conversion of British spellings to American spellings, (3) application of the built-in delete list to remove stop words (for each word deleted, a filler word, *xxx* is inserted in order to preserve the original spacing of the words in the text corpus), (4) application of a generalisation thesaurus to translate compound words into a single concept word, for example, converting *United States* or *United States of America* to *united_states* and (5) stemming, which involves identifying and standardising the root of a word so as to avoid multiple counts of a lexeme that has different forms but the same meaning (Lambert, 2017). For example, *live*, *lives* and *lived* were regarded as a single lexeme *live*. After performing Steps 1–5, a union concept-word list associated with each text corpus was generated. The two-union concept-word lists give the frequency of occurrence of each concept word derived from their corresponding text corpuses. To examine the quality of the text co-constructed by each group, each word on the two-union concept-word lists were coded as High (H) or Low (L) according to their lexical relevancy in the context of the two story analysis tasks. The two concept-word lists were coded by the first and second authors independently. The inter-rater reliability (Cohen's kappa) for the two coded union concept-word lists associated with Story 1 and Story 2 are 0.842 and 0.808 respectively, indicating that the coding was reasonably consistent. Thus, the number of concept words with high and low lexical relevancy in the text they co-constructed reflected the quality of students' work.

Text-processing phase

While the frequencies given in the concept-word list reflect, to a certain extent, the relative importance of the words used in a text corpus, they give little sense of the context of these concept words and how they connect with one another. In the text-processing phase, the key goal was to construct a co-occurrence semantic network, through which a deeper understanding of the contextual meaning of a text corpus could be made possible. This construction is grounded on the basic assumption that words that are located close to one another within a text are likely related in some ways (Lambert, 2017). In other words, the proximity between words is associated with the strength of their relationship. Building on this principle, AutoMap creates a window with a user-chosen window size for scanning through the text corpus sentence by sentence and word by word. Window size refers to the maximum distance between concept words to be connected. For instance, when applying a window size of 3 to the sentence "xxx story reveals xxx dark side xxx mankind xxx human wants xxx unlimited", 7 pairs of words are identified: story – reveals, reveals – dark, dark – side, side – mankind, mankind – human, human – wants, wants – unlimited. We chose a window size of 2 because using a lower number results in generating a more specific network (Carley et al., 2013). As a result, a semantic co-reference list (see Figure 1) was generated for each group report based upon the proximity of the concept words in a selected text or text fragment.

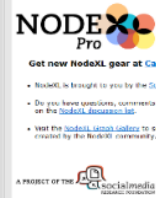<div align="center">(a)   (b)</div>

Figure 1. A semantic co-reference list generated by (a) AutoMap and imported into (b) NodeXL, where *Width* refers to the frequencies of occurrence of the word pairs

Data analysis phase

In this phase, we used NodeXL, a network analysis tool developed by the Social Media Research Foundation (Hansen et al., 2019), to unfold the complexity of each semantic network. To achieve this, we imported each semantic co-reference list associated with its corresponding group report to NodeXL, as shown in Figure 1 (b). The initial semantic networks derived from the co-reference lists are complex and often difficult to interpret as there are many overlapping vertices (concept words) and edges (connections between concept words). To gain a better sense of the structure of each semantic network, we conducted a cluster analysis with NodeXL by adopting the Clauset-Newman-Moore (2004) algorithm for grouping the vertices. The complexity of each semantic network can be discerned by counting the number of clusters and the number of vertices each cluster contains.

Different network measures such as degree centrality and betweenness centrality reflect the connectivity of each vertex. Centrality generally mirrors the relative importance of a vertex (concept word) in a semantic network. The degree centrality of a given vertex is defined as the total number of edges attached to it divided by the total number of vertices of the entire semantic network excluding itself. Thus, the degree centrality of a given vertex, to a certain extent, reflects the strength of its association with the rest of the semantic network. In comparison to degree centrality, betweenness centrality reflects the extent to which a vertex stands on paths linking one another. The betweenness centrality of Vertex V is defined as $g(V) = [(N-1)(N-2)/2]\sum_{S\neq V\neq T}\sigma_{ST}(V)/\sigma_{ST}$, where N is the total number of vertices of the semantic network, σ_{ST} is the total number of shortest paths linking Vertex S and Vertex T, while $\sigma_{ST}(V)$ is the number of those paths passing through V. An example of how betweenness centrality is computed is illustrated in Figure 2.
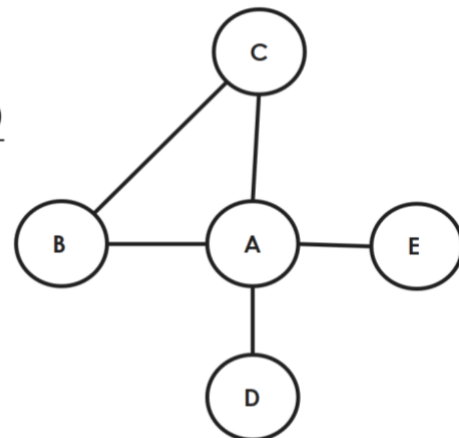

Figure 2. Computation of the betweenness centrality of a given vertex in a semantic network

In other words, betweenness centrality measures how in-between a vertex is among the other vertices in the semantic network. A vertex with a high betweenness centrality means that it serves as a hub that helps bridge several independent parts of the network. Thus, a semantic network with a high betweenness centrality indicates that the vertices or concept words within the semantic network are highly connected with one another. On the contrary, a low betweenness centrality implies that the semantic network may comprise a lot of isolated concept words or small disjoint networks (Traub et al., 2010). In short, the average betweenness centrality of a semantic network serves as a useful measure for gauging the connectivity among concept words and thus the quality of students' co-constructed text.

Results

To verify that the choice of two different stories had no statistically significant effect on the number of content words contributed by students, one-way between-group ANOVA and Mann-Whitney Test were conducted to compare the performance of the two independent sets of students – Set A and Set B – in using the discussion forum provided by Moodle to accomplish the story analysis task, where Set A and Set B were assigned to work on Story 1 and Story 2 respectively. As shown in Table 2, the *p* values given by the one-way between-group ANOVA and Mann-Whitney test are 0.266 and 0.351 respectively. These results indicate that the effect of the choice of story is statistically insignificant to the number of content words contributed by individual students.

Table 2
One-way between-group ANOVA of the number of content words contributed by individual students from Set A and Set B through the discussion forum

Story 1 (Set A)		Story 2 (Set B)		One-way between-group ANOVA		Mann-Whitney test	
Mean	SD	Mean	SD	F(1, 25)	*p* value	Mann-Whitney U	*p* value
102.8	57.4	133.5	80.0	1.296	0.266	71.0	0.351

Students' participation in the two online settings

To compare students' participation in Moodle's forum with that in Diigo coupled with Google Docs, the number of postings contributed separately by students of Set A and Set B is plotted against time. Figure 3 illustrates the variation in the number of postings per day contributed the two sets of students. For both story analysis tasks, the number of postings is apparently higher in the set of students using Diigo coupled with Google Docs than in those using Moodle's forum. Likewise, the activity span of the set of students using Diigo coupled with Google Docs to accomplish the tasks is longer than those using Moodle's forum.
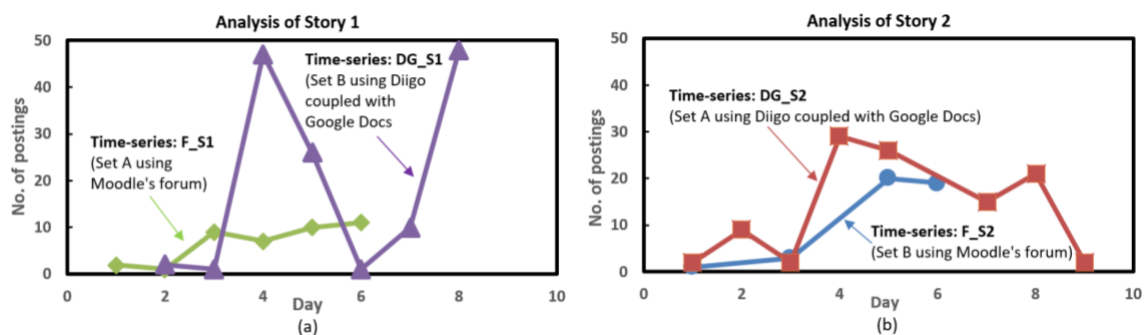

Figure 3. Variation of the number of postings contributed by students over time for the task associated with Story 1 (a) and Story 2 (b)

To compare students' activity patterns in the two online settings, the cross-correlation analysis was conducted between the four time series – F_S1, DG_S1, F_S2 and DG_S2, shown in Figure 3 – using IBM SPSS version 25 package. The cross-correlation function enables us to compute the correlation between the observations of two time series x_t and y_t, lagged by h time units (i.e., the correlation between y_{t+h} and x_t), where h can be varied from -3 to 3. Figure 4 (a) shows that all cross-correlation function values for different lag time (h) units fall below the confidence limits, indicating that there is no statistically significant

correlation between the two time series F_S1 and DG_S1. Similarly, Figure 4 (b) also indicates that F_S2 and DG_S2 do not correlate with one another. In other words, students' activity patterns found in the Diigo coupled with Google Docs setting are statistically different from those found in the Moodle's forum setting. Thus, using Diigo coupled Google Docs to support collaborative work seems to be effective in enhancing student–student interaction and extending their activity spans.
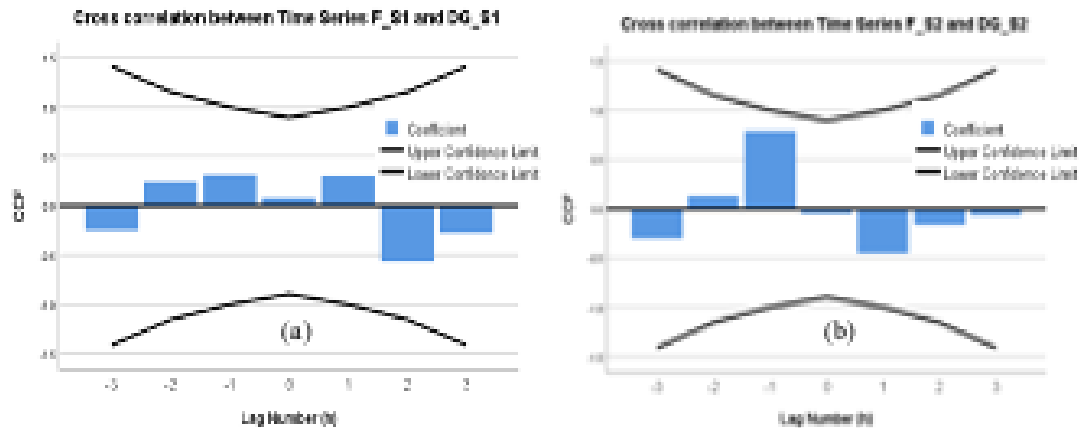

Figure 4. Cross-correlation between the time series of students' activities on (a) Story 1 using Moodle's forum (F_S1) and Diigo coupled Google Docs (DG_S1); and (b) Story 2 using Moodle's forum (F_S2) and Diigo coupled with Google Docs (DG_S2)

Comparison of student engagement in the two online settings

To compare the effects of using Diigo coupled with Google Docs with that of using Moodle's forum on student engagement in collaborative learning, the content words contributed by each individual student were counted. As given in Table 3, for the two story analysis tasks, the results of the one-way between-group ANOVA show that the average number of content words contributed by students using Diigo coupled with Google Docs is significantly greater than that of those using Moodle's forum. For the task associated with Story 1, the effect size (η^2) is 0.15, which is a large effect according to commonly used guidelines proposed by Cohen (1988) (~0.01 = small effect, ~0.06 = moderate effect, $\geq$ 0.14 = large effect), indicating that the intervention of using Diigo and Google Docs accounts for 15% of the variance of the number of content words contributed by individual students. For Story 2, the effect size (η^2) is 0.21, indicating again that the intervention exhibits a large effect on number of content words contributed by individual students and accounts 21% of its variance.

Table 3
One-way between-group ANOVA of the impact of the use of the collaborative tools on the number of content words contributed by individual students from Set A & Set B

	Forum		Diigo & Google Docs				
	Mean	SD	Mean	SD	F(1, 25)	*p* value	Effect size (η^2)
Story 1	102.8[a]	61.6[a]	155.5[b]	73.4[b]	4.270	0.049	0.15
Story 2	133.5[b]	79.9[b]	276.8[a]	194.3[a]	6.459	0.018	0.21

Note. [a] Results associated with Set A; [b] results associated with Set B.

Since each student was engaged in two trials of story analysis activities, one with Moodle's forum and the other with Diigo coupled with Google Docs, we compared their word contribution between the two trials by using one-way repeated measure ANOVA. The results in Table 4 indicate that the number of content words contributed by individual students in using Diigo coupled with Google Docs is significantly greater than that using Moodle's forum and that η^2 is 0.27, suggesting a large effect size which accounts for 27% of the within-group variance.

Table 4
One-way repeated measures ANOVA for comparison of content words contributed by individual students in the two trials

Forum		Trial Diigo & Google Docs		Wilks' lambda	$F_{(1, 26)}$	*p* value	Effect size (η^2)
Mean	SD	Mean	SD				
118.7	70.5	213.9	154.7	0.73	9.604	0.005	0.27

Semantic network analysis

To discern the underlying structure of each network, a cluster analysis was conducted by using the cluster analytical tool provided in NodeXL. Each cluster of vertices (concept words) displayed in the visualisation was grouped together by the Clauset-Newman-Moore (1984) cluster algorithm, based on the co-occurrence frequency of each vertex pair. In a nutshell, the vertices within each cluster co-occur with one another more frequently than they do with other vertices. According to Drieger (2013), each cluster can be interpreted as a strongly connected component encoding a specific semantic topic or complex concept. Thus, the number of clusters identified reflects the semantic richness of its associated text. To illustrate the clustering of vertices, the semantic co-reference lists derived from groups of the same set were merged to form a joint semantic network and clusters were identified with Clauset-Newman-Moore algorithm (see Figure 5). The grey lines and grey curves represent respectively the intra- and inter-cluster connections. The thickness of each line and curve corresponds to the strength of each connection. These networks of clusters provide an overall sense of the complexity and semantic richness of the text co-constructed by students in the two online settings.
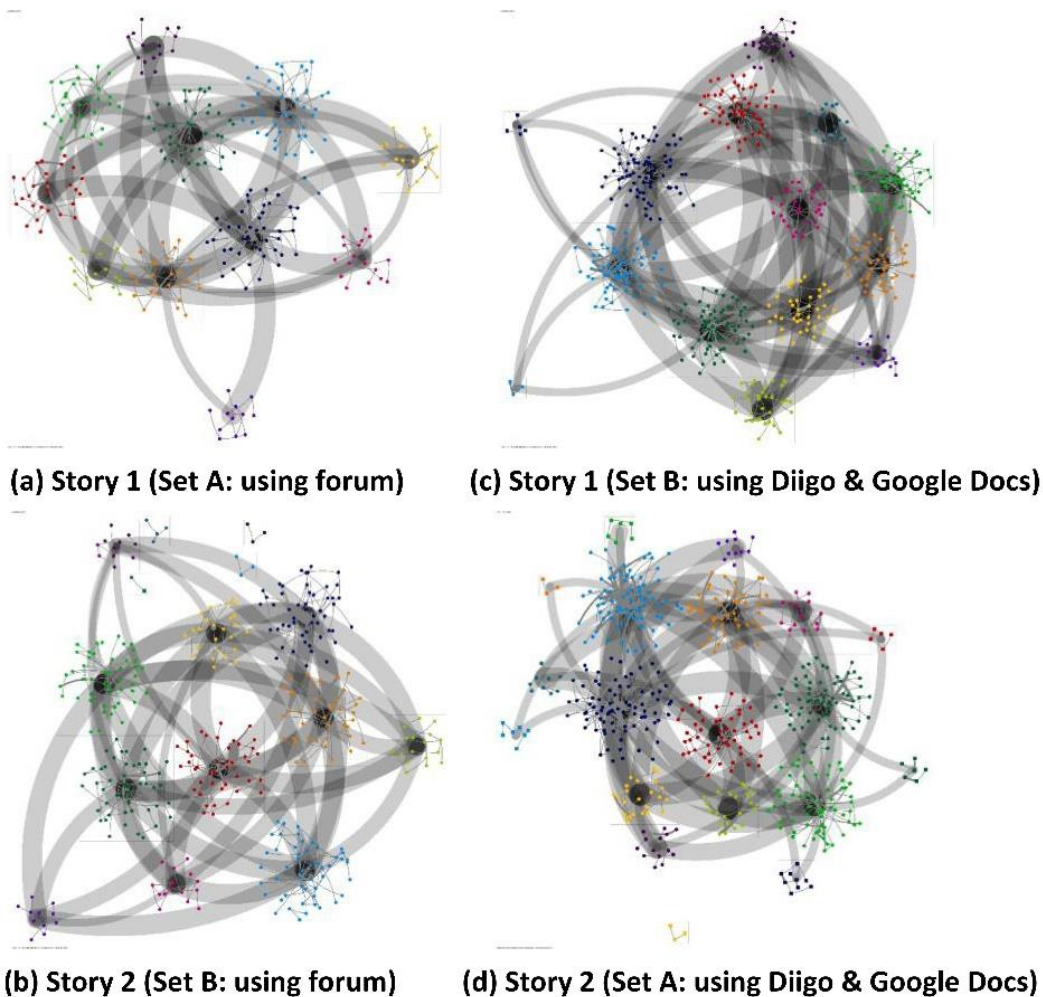

(a) Story 1 (Set A: using forum) **(c) Story 1 (Set B: using Diigo & Google Docs)**

(b) Story 2 (Set B: using forum) **(d) Story 2 (Set A: using Diigo & Google Docs)**

Figure 5. Merged semantic networks generated from the text contributed by students from Set A and Set B

The number of vertices and the average values of the betweenness centrality of the semantic networks identified in story analysis conducted in the two online settings are listed in Table 5. The number of clusters identified in the semantic networks associated with each group in the two online settings is given in the first two columns of Table 5. To reduce the noises in the dataset, only clusters with a size of 5 or more vertices were considered for analysis (Kok & Domingos, 2008). The results depicted in Table 5 indicate that the number of clusters and vertices and the average betweenness centrality identified in the story analysis conducted in the setting using Diigo coupled with Google Docs is generally larger than the corresponding semantic measures identified in the story analysis using Moodle's forum. The last two columns give the average grade score of each group's story analysis assessed independently with an intra-class coefficient (ICC) is 0.882 for single measures and 0.937 for average measures, indicating that the inter-rater reliability is high.

Table 5
Semantic network measures of the summary text of the story analysis conducted by each group in the two online settings

	Clusters		Vertices		Average betweenness centrality		Average grade score	
Group	F	DG	F	DG	F	DG	F	DG
A1	7	11	112	842	145.5	265.8	11.5	39.5
A2	10	12	742	642	274.9	281.3	35.0	38.0
A3	8	12	174	1064	121.3	543.6	13.0	42.0
B1	9	12	512	572	241.2	248.8	23.0	30.5
B2	10	11	528	680	263.9	277.7	24.0	39.5
B3	10	16	458	962	264.0	525.0	26.5	39.5

Note. F refers to trial using Moodle's forum, while DG refers to the trial using Diigo coupled with Google Docs.

The Pearson correlation coefficient of the average betweenness centrality and the average grade score is 0.742 with a *p* value of 0.006, indicating that the two quantities correlate strongly with one another. To compare the semantic network measures identified in the two online settings, parametric methods such as paired-samples *t* test or repeated measures ANOVA may not be an appropriate choice for analysis as the sample size is small and normality assumption would be difficult to uphold. To circumvent this situation, the Wilcoxon Signed Rank Test (Pallant, 2007), a non-parametric counterpart of one-way repeated measures ANOVA, was employed to compare the semantic network measures associated with the two online settings. As depicted in Table 6, the numbers of clusters and the average betweenness centrality identified in the story analysis associated with the use of Diigo coupled with Google Docs are higher than those associated with Moodle's forum at significant levels of 0.027 and 0.028 respectively. On the other hand, the numbers of vertices found in the semantic networks associated with Diigo coupled with Google Docs are generally higher than those associated with Moodle's forum, their differences are marginally significant, at a level of 0.075. For Group A2, the number of vertices associated with Moodle's forum is even larger than that associated with Diigo coupled with Google Docs.

Table 6
Wilcoxon Signed Rank Test for comparing the semantic network measures associated with the story analysis conducted in Moodle's forum (F) and Diigo coupled with Google Docs (DG)

	Clusters	Vertices	Average betweenness centrality
	F vs DG	F vs DG	F vs DG
Z statistics	-2.21	-1.78	-2.20
Asymptotic significant level	0.027	0.075	0.028

To further scrutinise the quality of each vertex, the two union concept-word lists associated with Story 1 and Story 2 were coded according to their lexical relevancy to the context of the two respective story analysis tasks. So, each vertex was labelled with high (H) and low (L) lexical relevancy. The numbers of vertices with high and low lexical relevancy were listed in Table 7. Results of the Wilcoxon Signed Rank Test given in Table 8 show that the average number of vertices with high lexical relevancy associated with Diigo coupled with Google Docs is significantly larger than that associated with Moodle's forum (*p* value ~0.028).

Table 7
The numbers of vertices with a high and low lexical relevancy associated with the text co-constructed by each group using Moodle's forum (F) and Diigo coupled with Google Docs (DG)

	No. of vertices with high lexical relevancy		No. of vertices with low lexical relevancy	
Group	F	DG	F	DG
A1	2	44	110	798
A2	28	34	714	608
A3	12	123	162	941
B1	23	38	489	534
B2	21	40	507	640
B3	18	32	440	930

Table 8
Wilcoxon Signed Rank Test for comparing vertices with different lexical relevancy associated with the story analyses conducted in Moodle's forum (F) and Diigo coupled with Google Docs (DG)

	No. of vertices with high lexical relevancy per person (LR_H)	No. of vertices with low lexical relevancy per person (LR_L)
	F vs DG	F vs DG
Z statistics	-2.20	-1.78
Asymptotic significant level	0.028	0.075

Level of collaboration in Diigo and Google Docs

To examine further how group collaboration in the online setting associated with Diigo coupled with Google Docs affects the quality of students' collaborative work, we adopted the metric proposed by Li et al. (2015) to gauge the level of collaboration within a group. *Collab* is defined as the number of posted items divided by the normalised standard deviation (σ') of the number of posted items as shown in Table 9.

Table 9
The mean and standard deviation of each group's postings in Diigo coupled with Google Docs

Group	No. of postings (n)	Av. no. of postings per person (μ)	σ	σ'	Collab
A1	17	3.4	3.78	1.11	15.28
A2	40	9.6	7.96	0.83	57.92
A3	79	15.8	6.30	0.40	198.10
B1	64	14.75	11.73	0.80	74.19
B2	9	2.25	1.89	0.84	10.70
B3	60	15	10.86	0.72	82.85

Note. σ' is defined as σ/μ; Collab is defined as n/σ'

Although μ reveals the intensity of each group's activities in Diigo coupled with Google Docs, σ' reflects if each member took a fair share of contribution to the assigned group task. In other words, a small σ' value indicates that each group member contributed equally well to the group work. Thus, a high score of Collab implies that the group activities are intensive (reflected by having a large μ) and that each group member has a fair share of contribution to the group task (reflected by having a small σ').

As presented in previous sections, while the concept connectivity of each semantic network is associated with its betweenness centrality, the quality of the text co-constructed by each group can be reflected by the number of vertices with different levels of lexical relevancy. To examine how the level of collaboration with a group affected the concept connectivity and lexical relevancy, correlations between Collab and the average betweenness centrality and between Collab and numbers of vertices with different levels of lexical relevancy associated with the text co-constructed by each group were computed. The results given in Table 10 indicate that the level of collaboration, Collab correlates strongly with the average betweenness centrality (BT) and the number of vertices with high lexical relevancy (LR_H), with respective correlation coefficients of 0.77 and 0.85. On the other hand, the correlation between Collab and LR_L is statistically

insignificant. These indicate that collaboration within a group helps enhance the semantic richness as well as the quality of the artefact co-constructed by each group.

Table 10
Correlation of collaboration index with selected semantic network measures and number of vertices with different lexical relevancy levels

	Pair of variables	Pearson correlation coefficient	*p* value (one-tailed)
Collaboration index (Collab)	Av. betweenness centrality (BT)	0.77	0.036
	No. of vertices with high lexical relevancy (LR_H)	0.85	0.016
	No. of vertices with low lexical relevancy (LR_L)	0.54	0.136

Discussion and conclusion

The above results suggest that the pedagogical use of Diigo coupled with Google Docs has a positive impact on enhancing student engagement and the quality of their contribution to the collaborative work. The time-series analysis reveals that using Diigo coupled with Google Docs to support collaborative work seems to be more effective in enhancing student engagement and extending their activity spans. Students generally posted more notes and remained engaged in activities for a longer duration in the DG setting. This finding corroborates with previous studies on students' perceived learning experience on social annotation (Gao, 2013; Miettinen et al., 2003; Nokelainen et al., 2005) and online collaborative writing (M. Liu et al., 2018; Zhou et al., 2012). In comparison to the work conducted by Johnson et al. (2010), our study provides rich empirical data of student learning process to support the claim that Diigo coupled with Google Docs can be used to sustain collaboration over a prolonged period of time. Our interpretation is that social annotation enables students to granulate the entire discourse within the group into more confined but distributed discussions anchored at specific sections or sentences or words appeared in the text. Students can break down an issue into sub-tasks to make the discussions more focused, elaborated and manageable. The anchored notes posted by students at specific locations of the text helps direct other members' attention to important issues and provide peer scaffolding for others to build on (Gao, 2013). Students can leave their spontaneous thoughts or reflections there and then while reading. Diigo coupled with Google Docs seems to constitute a better environment for fostering student interaction and learning engagement as indicated by the intensity of students' activities and number of content words contributed by each group. We argue that while Diigo provides a mechanism that favours a less structured and divergent discourse, Google Docs offers a conducive environment for students to collate and re-organise their thoughts developed through social annotation and to further elaborate on ideas and extend their writing collaboratively; and that the high visibility of collaborators' digital traces in Google Docs and its affordance for simultaneous co-construction may incentivise students to engage in the collaborative process. So, coupling social annotation with collaborative writing facilities can help alleviate students' cognitive load in managing their notes and extend their thoughts beyond their annotations. Indubitably, to substantiate this claim, a more in-depth analysis of students' learning experiences with Diigo coupled with Google Docs is deemed necessary.

Apart from enhancing students' engagement in and contribution to the collaborative tasks, the use of Diigo coupled with Google Docs also exerted positive impacts on the quality of work delivered by each group. All the semantic measures such as the number and size of semantic clusters, average betweenness centrality and lexical relevancy, were higher for those groups working in the Diigo coupled with Google Docs setting. These semantic measures help unfold the complexity and quality of the work accomplished by each group. While average betweenness centrality reflects the connectivity of the vertices in the network, the number and size of the clusters mirror the semantic richness of the associated text and lexical relevancy reflects the depth of the domain knowledge exhibited in students' work. Researchers may question the validity of using these measures to assess students' work. However, the strong correlation between the average grade score and the average betweenness centrality associated with each group's story analysis report indicates that the average betweenness centrality is indeed a good measure for gauging the quality of students' work.

Another interesting finding is that, for those groups working in the Diigo coupled with Google Docs setting, their level of collaboration correlates strongly with the average betweenness centrality of the semantic network and the number of vertices with high lexical relevancy associated with their co-constructed text. It suggests that the level of collaboration predicts the quality of students' collaborative work. The relationship

between collaboration and learning outcomes has been well deliberated in literature. Nonetheless, very few studies employed quantitative measures to assess students' group processes and their learning outcomes. Our findings provide insights into the development of appropriate objective measures for assessing students' collaborative processes and the quality of their work.

In sum, the results of this study provide empirical evidence that augmenting social annotation with collaborative writing facilities enhances student engagement in the collaborative process and the quality of student work as reflected by the semantic richness, concept connectivity and lexical relevancy of the text co-constructed by each group. The findings also suggest that students' active participation in the collaborative process helps promote the level of their cognitive processes and the quality of the learning outcomes. Diigo coupled with Google Docs seems to provide the essential educational affordance for students to engage in in-depth discussions, meaning negotiation and knowledge co-construction. The findings help inform future development of a more integrated social annotation environment. From a methodological point of view, our findings demonstrate the usefulness of semantic network analysis in providing relevant quantitative measures for gauging the semantic richness, concept connectivity and lexical relevancy of the text co-constructed by each group. Furthermore, the current study also contributes to the growing body of research on learning analytics and adaptive learning environments. The quantitative measures adopted in our study provide insights into the development of useful learning analytics for assessing and predicting students' online learning and designing appropriate adaptive support to students.

Limitations and future research

Given the small sample size chosen in this study, we have no intention of making any generalised claims on the affordance of social annotation coupled with online collaborative writing in facilitating knowledge co-construction. The findings should be construed as illustrative rather than conclusive. The aim of this study was to provide insights into designing collaborative learning activities with Diigo coupled with Google Docs and the adoption of appropriate quantitative measures in delineating students' learning trajectories in online environments. Although the use of Diigo coupled with Google Docs seems to have a positive effect on collaborative learning, little is known about how pedagogical factors, such as task design and teacher's facilitation, shape students' cognitive processes and level of collaboration. Thus, close scrutiny of dynamic interaction between students and teachers is necessary in future studies. Furthermore, apart from betweenness centrality, it is worth exploring the sensitivity of other semantic measures such as eigenvector centrality or page rank centrality in assessing the quality of students' co-constructed artefacts.

Acknowledgements

References

Burke, A. (2011). Group work: How to use groups effectively. *The Journal of Effective Teaching, 11*(2), 87–95. https://uncw.edu/jet/articles/vol11_2/burke.pdf

Carley, K. M., Columbus, D., & Landwehr, P. (2013). *Automap user's guide 2013*. Carnegie Mellon University. http://www.casos.cs.cmu.edu/projects/automap/CMU-ISR-13-105.pdf

Chu, S. K. W., Zhang, Y., Chen, K., Chan, C. K., Lee, C. W. Y., Zou, E., & Lau, W. (2017). The effectiveness of wikis for project-based learning in different disciplines in higher education. *Internet and Higher Education, 33*, 49–60. https://doi.org/10.1016/j.iheduc.2017.01.005

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E, 70*(6), Article 066111. https://doi.org/10.1103/PhysRevE.70.066111

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.

Dean, J. (2015, August 25). Back to school with annotation: 10 ways to annotate with students. *Hypothes.is*. https://web.hypothes.is/blog/back-to-school-with-annotation-10-ways-to-annotate-with-students/

Drieger, P. (2013). Semantic network analysis as a method for visual text analytics. *Procedia – Social and Behavioral Sciences, 79*, 4–17. https://doi.org/10.1016/j.sbspro.2013.05.053

Feinerer, I., & Hornik, K. (2020). *Package 'tm'*. The CRAN Repository. https://cran.r-project.org/web/packages/tm/tm.pdf

Gao, F. (2013). A case study of using a social annotation tool to support collaboratively learning. *The Internet and Higher Education, 17*, 76–83. https://doi.org/10.1016/j.iheduc.2012.11.002

Hansen, D., Himelboim, I., Shneiderman B., & Smith, M. A. (2019). *Analyzing social media networks with NodeXL: Insights from a connected world* (2nd ed.). Morgan Kaufmann.

Haya, P. A., Daems, O., Malzahn, N., Castellanos, J., & Hoppe, H. U. (2015). Analysing content and patterns of interaction for improving the learning design of networked learning environments. *British Journal of Educational Technology, 46*(2), 300–316. https://doi.org/10.1111/bjet.12264

Hoser, B., Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006). Semantic network analysis of ontologies. In Y. Sure & J. Dominque (Eds.), *Lecture notes in computer science: Vol. 4011. The semantic web: Research and applications* (pp. 514–529). Springer. https://doi.org/10.1007/11762256_38

Johnson, T. E., Archibald, T. N., & Tenenbaum, G. (2010). Individual and team annotation effects on students' reading comprehension, critical thinking, and meta-cognitive skills. *Computers in Human Behavior, 26*, 1496–1507. https://doi.org/10.1016/j.chb.2010.05.014

Kear, K., Jones, A., Holden, G., & Curcher, M. (2016). Social technologies for online learning: Theoretical and contextual issues. *Open Learning, 31*(1), 42–53. https://doi.org/10.1080/02680513.2016.1140570

Kok, S., & Domingos, P. (2008). Extracting semantic networks from text via relational clustering. In W. Daelemans, B. Goethals, & K. Morik (Eds.), *Lecture notes in computer science: Vol. 5211. Machine learning and knowledge discovery in databases* (pp. 624–639). Springer. https://doi.org/10.1007/978-3-540-87479-9_59

Lambert, N. J. (2017). Text mining tutorial. In A. Pilny & M. S. Poole (Eds.), *Group processes: Data-driven computational approaches* (pp. 93–117). Springer. https://doi.org/10.1007/978-3-319-48941-4_5

Li, S. C., Pow, J. W. C., & Cheung, W. C. (2015). A delineation of the cognitive processes manifested in a social annotation environment. *Journal of Computer Assisted Learning, 31*(1), 1–13. https://doi.org/10.1111/jcal.12073

Liu, M., Liu, L., & Liu, L. (2018). Group awareness increases student engagement in online collaborative writing. *The Internet and Higher Education, 38*, 1–8. https://doi.org/10.1016/j.iheduc.2018.04.001

Liu, S. H.-J., & Lan, Y.-J. (2016). Social constructivist approach to web-based EFL learning: Collaboration, motivation, and perception on the use of Google Docs. *Educational Technology & Society, 19*(1), 171–186. https://drive.google.com/open?id=1lFXir9paiCM_WgraMXU55BeD_3C5TvIZ

Mendenhall, A., & Johnson, T. E. (2010). Fostering the development of critical thinking skills, and reading comprehension of undergraduates using a Web 2.0 tool coupled with a learning system. *Interactive Learning Environments, 18*(3), 263–276. https://doi.org/10.1080/10494820.2010.500537

Miettinen, M., Nokelainen, P., Flor´een, P., Tirri, H., & Kurhila, J. (2003). EDUCOSM - personalized writable web for learning communities. In P. K. Srimani, W. Bein, R. Hashemi, E. Lawrence, M. Cannataro, E. Regentova, & A. Spink (Eds.), *Proceedings of the International Conference on Information Technology: Coding and Computing* (pp. 37–42). IEEE. http://doi.org/10.1109/ITCC.2003.1197496

Neumann, D. L., & Hood, M. (2009). The effects of using a wiki on student engagement and learning of report writing skills in a university statistics course. *Australasian Journal of Educational Technology, 25*(3), 382–398. https://doi.org/10.14742/ajet.1141

Nokelainen, P., Miettinen, M., Kurhila, J., Floréen, P., & Tirri, H. (2005). A shared document-based annotation tool to support learner-centred collaborative learning. *British Journal of Educational Technology, 36*(5), 757–770. https://doi.org/10.1111/j.1467-8535.2005.00474.x

Olenewa, R., Olson, G. M., Olson, J. S., & Russell, D. M. (2017). Now that we can write simultaneously, how do we use that to our advantage? *Communications of the ACM, 60*(8), 36–43. https://doi.org/10.1145/2983527

Pallant, J. (2007). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows* (3rd ed.). Open University Press.

Seyyedrezaie, Z. S., Ghonsooly, B., Shahriari, H., & Fatemi, H. H. (2016). A mixed methods analysis of the effect of Google Docs environment on EFL learners' writing performance and causal attributions for success and failure. *The Turkish Online Journal of Distance Education, 17*(3), 90–110. https://dergipark.org.tr/en/download/article-file/222623

Suwantarathip, O., & Wichadee, S. (2014). The effects of collaborative writing activity using google Docs on students' writing abilities. *The Turkish Online Journal of Educational Technology, 13*(2), 148–156. http://www.tojet.net/articles/v13i2/13215.pdf

Traub, M. C., Walter, W., & Lamers, M. H. (2010). A semantic centrality measure for finding the most trustworthy account. In H. Weghorn, J. Roth, & P. Isaias (Eds.), *Proceedings of the IADIS International Conference Informatics* (pp. 117–125). IADIS. http://www.iadisportal.org/digital-library/a-semantic-centrality-measure-for-finding-the-most-trustworthy-account

Zhou, W., Simpson, E., & Domizi, D. P. (2012). Google Docs in an out-of-class collaborative writing activity. *International Journal of Teaching and Learning in Higher Education, 24*(3), 359–375. https://www.isetl.org/ijtlhe/pdf/IJTLHE24(3).pdf

---

Corresponding author: Sandy C. Li, sandyli@hkbu.edu.hk