Idea Identification and Analysis (I²A): A Search for Sustainable Promising Ideas Within Knowledge-Building Discourse

Vwen Yen Alwyn Lee, Nanyang Technological University, alwynlee@ntu.edu.sg
Seng Chee Tan, Nanyang Technological University, sengchee.tan@ntu.edu.sg
Joon Kit Kelvin Chee, Ministry of Education, chee_joon_kit_kelvin@moe.edu.sg

Abstract: This is a study on the identification and analysis of promising ideas, which examines the knowledge building discourse of 20 students within Knowledge Forum, through the I²A system in a mixed method approach. It is a challenge in providing support to maintain efforts in identifying and improving ideas, which this study addresses through visualization and step-wise analysis of discourse network measures that are calculated using the Knowledge Building Discourse Explorer. By using Social Network Analysis with chosen keywords and analyzing quality of notes, we are able to identify and trace the evolution of ideas that become promising talking points within the discourse space. Findings show that ideas can be classified and further analyzed to have an impact on subsequent discourse. Teachers can focus their resources on promising ideas that help the community maintain engagement in creative work to improve ideas, so as to learn more effectively and attain knowledge creation goals.

Introduction

In educational institutions, learning approaches have long been operating within the acquisition and participation metaphors (Sfard, 1998), but there is also a new approach which is seen as an integration of both metaphors of learning (Paavola, Lipponen & Hakkarainen, 2004). Sfard (1998) examined the acquisition and participatory metaphors of learning and suggested that both cognitive and situative perspectives are required. Paavola et al. (2004) further suggested that the combination of both metaphors could be problematic, but an integration of both would require a third approach: the knowledge-creation metaphor of learning. Scardamalia and Bereiter's model of knowledge creation, termed Knowledge Building (Scardamalia & Bereiter, 2003), is one of the three influential models reviewed by Paavola et al. (2004), besides transformation of activity system (Engeström, 1999) and creation of new products at organizational level (Nonaka & Takeuchi, 1995). Knowledge building focuses on K-12 learning environment that leverages learner's natural idea generation capability to work on continual improvement of their ideas, and for teachers to maintain student engagement in idea improvement. Scardamalia and Bereiter (2003) distinguished between two modes of learning: namely, the belief mode and the design mode. The belief mode engages learners with what they or others believe or ought to believe, whereas the design mode engages learners with the usefulness, adequacy, improvability and developmental potential of ideas. Knowledge building as a social process makes use of both learning modes, where learners engage in collaborative inquiry to enable creation, contribution and advancement of community knowledge. The concept of idea improvement (Scardamalia, 2002) is crucial as students are required to have motivation in their search for solutions to handle significant challenges, be aware of emergent themes of inquiry from multiple sources of inputs, acknowledge knowledge gaps, and participate in collaborative idea improvement (Zhang, Scardamalia, Reeve & Messina, 2009).

In order to encourage and promote continuous efforts in idea improvement, we need a process of sharing and improvement of ideas within the community, and also to identify potential relevant ideas that show relative significance and promise to the community, such that these promising ideas could be brought forward and be continuously improved upon. Isolated suggestions and ideas from individuals are often unable to frame the problem accurately and cannot provide a complete picture of the ideal solution in most problems. However, students might not have the capability to continuously maintain engagement in the process of idea improvement over long periods of time. In such situations, teachers need to provide relevant support and scaffolds, but reading and monitoring student discourse can be challenging, especially when students' posts cumulate and become increasingly unmanageable over time. With appropriate tools and technology, such processes can be facilitated for greater efficiency and effectiveness. For example, technology and tools can be used for visualization and analysis, to automate the process of identifying promising ideas to allow efficient learning of content, and to also provide the ability to track the impact after the introduction of promising ideas into the discourse community. If this can be achieved, students would be able to develop ideas to attain knowledge creation goals within the classroom, and they would not be engaged in other peripheral and less important ideas, which could sideline and affect their immersion in progressive problem solving (Bereiter & Scardamalia, 1993). Identifying promising ideas within a discourse is a reasonable and natural approach towards efficient learning, as discourse often plays a creative role in encouraging improvements on ideas (Lakatos, 1970). Platforms such as Knowledge Forum

(Scardamalia, 2004) and software tools like Knowledge Building Discourse Explorer (KBDeX) (Oshima, Oshima, & Matsuzawa, 2012) are used for supporting knowledge building pedagogies, practices and research. These platforms and tools help to characterize constructive interaction between students, not limited to just social interactions but also allowing discourse to play a part in determining the nature of ideas that surfaced during discourse. By doing step-wise analysis of idea occurrences and improvements brought about by different discourse units, the flow and content of ideas can be predicted to improve over the discourse, which can be used as feedback to inform future classes and for contributing back to the community.

In this paper, we seek to identify relevant and promising ideas which may not be immediately apparent to teachers and students, and propose an automated method of tagging promising ideas, which was previously done through manual consensus by students (Chen, Chuy, Resendes, & Scardamalia, 2010). The proposed method of tagging promising ideas and analysis of subsequent discourse would help in deciding whether an idea has surfaced within the discourse, and if it has sufficient promise and potential to warrant timely interventions so as to sustain creative work in improving emergent promising ideas that eventually lead to community "rise-above" and more effective learning. Rise-above refers to the integration of ideas or elevation of explanation at a higher theoretical or principle level. The method is conducted in the following phases: (a) identification of non-obvious ideas among the large number of threads and chatter within online discourse, which might be missed by teachers or students; (b) determination of promising ideas that can sustain engagement of students and contribute to the community discourse; (c) evaluation of the impact and contribution of promising ideas to the community such that timely interventions could provide sustained students' engagement in creative work to improve ideas.

The focus on learning process, discourse units and epistemic words

In the world of knowledge acquisition, designers construct methods to allow learners to attain pre-determined learning goals; whereas in the world of learning, designers focus on methods that support learning processes, rather than the acquisition of knowledge and pre-defined learning goals (Strijbos, Martens & Jochems, 2004). It is crucial to focus on the learning process in which students seem to benefit from, even though pre-planned outcomes might deviate. The learning process within online discourse is relevant to the concept of continuously improving and sharing of students' ideas for the community to rise above, and this process is not just limited to the social interaction of students, but also applicable to the usage of discourse units (known as notes within Knowledge Forum) and words within the learning environment. The ability to identify promising ideas, ideas that become consequential when worked on, is essential for creative work with ideas at all levels (Chen, Scardamalia, & Resendes, 2013). This leads to the need to analyze text within online discourse, which are representations of opinions and views of students in the digital format, comparable to speech and presentations within traditional classroom discourse in the physical and audible form. There are various forms of discourse analysis, including automated content analysis that uses Natural Language Processing and Machine Learning techniques (Sun, Zhang, Jin, & Lyu, 2014) to focus on the appealing properties of unsupervised learning in topical models. The objective of probabilistic topic models is to automatically discover topics within a corpus and topic assignments of documents based on frequency of unique words. As compared to common choices of topic models such as latent semantic indexing (LSI) (Hofmann, 2001) and latent Dirichlet allocation (LDA) (Blei, 2012), this research focuses on unique words similar to Sun et al. (2014). In addition, we explore deeper than the concept of topics and themes by identifying the specific ideas that the writer is trying to portray through the linkage and usage of words at different temporal positions within the discourse.

Further, the analysis of discourse units and words reflect individuals' contributions to community knowledge, which cannot be analyzed through pre- and post-tests (Oshima et al., 2012), while the knowledge levels that are attained by students are commonly investigated through activity coding such as the coding scheme conducted by van Aalst (2009) on secondary students' written discourse. Apart from the common practice of using Social network Analysis (SNA) in analyzing social patterns of learners, this study focuses on emergent development patterns of ideas. By using SNA based on words that learners use in their discourse, ideas can be identified and the level of community knowledge obtained by learners could be represented in a novel way to reflect promising potential of emergent ideas at a point in time within the discourse. SNA could be used as a new representation of community knowledge building by learners (Oshima et al., 2012), which researchers can use to further investigate different models associated with the knowledge building community. KBDeX as a software tool was utilized in this research as a graphical visualization for identified ideas, and through our analysis, was used to support interpretations of promising ideas within a community discourse.

Methods

The proposed I²A system in this paper utilized knowledge building platform Knowledge Forum (Scardamalia & Bereiter, 2006) for conducting online discourse. KBDeX (Oshima et al., 2012) was used for analyzing knowledge

building discourse using social network measures, so as to identify ideas with their degree of promisingness which have a subsequent impact on community discourse. The discovery of promising ideas and their effects through social network analysis were subsequently qualitatively validated by analyzing the quality of notes.

Discourse platform and context

Using the Knowledge Forum as a collaborative knowledge building environment, this research analyzes the online discourse of 20 eighth-graders who were being taught the topic of "Human Transport System" over a period of two weeks. We worked closely with a middle school science teacher, who was able to plan and execute the knowledge building lesson for the students in a computer-aided environment. The teacher involved within the study also had prior knowledge building training and was able to facilitate the knowledge building lessons effectively, having previously used physical "idea cards" to trigger interactions and present authentic problems. Students had experience using Knowledge Forum and were keen to continue using it, as they found that the social interactions could help advance community knowledge. Each student was provided with a booklet on a fictional character "Uncle Yong", who has an impending heart attack. This problem was authentic to some students as their family members had previously encountered such an issue, thus the problem acted as a trigger aimed at eliciting students' ideas about heart problems. A motivated student (Student S2) posed four thought-invoking questions that were closely related to issues that his peers would comprehend. These questions were: *What heart problem is he facing? How can he improve on his diet and his lifestyle to get healthier? How will the operation be done? What will be the consequences if he continues to eat unhealthily?* Responses were posted as notes within Knowledge Forum on the "Human Transport System" view, which served as a workspace for all written notes within the two weeks learning session. The questions encouraged discussions and debates among students, with the usage of scaffolds such as "I need to understand" when posting notes of query, "My Theory" and "New Information" for statements of information that they have garnered online.

Archiving ideas and offline discussions into online notes on Knowledge Forum allow teachers and researchers to retrieve, assess and analyze the discourse and idea generation process. The discourse consist of 101 notes shared by 20 students, where all students contributed at least one note of significant content. Some students were very active within Knowledge Forum, constantly replying to other peers' contributions and were responsible for the majority of build-on and rise-above notes. The text within each note in Knowledge Forum were imported into KBDeX for visualization and analysis, and the majority of these notes were individual expressions of one or more ideas, apart from the rise-above notes which were co-authored by multiple students and represented collective ideas. Every note that represented expression of ideas and used within KBDeX, would be termed as a "discourse unit", or DU for short, represented visually as nodes in a network. Examples of such discourse units containing their respective authors and text are shown in Figure 1.

Name	Text
9	my theory - his blood vessels in the heart were damaged due to the blockage of fatty deposits, he finds it hard to breathe and he is having chest
10	i need to understand - is the plaque blocking the coronary artery the same as the plaque in your teeth?-

Figure 1. Excerpts of discourse units (DUs) posted by students 9 and 10.

Knowledge building discourse explorer (KBDeX) and BC trends

KBDeX was developed as a cohesive and easy-to-use platform that affords visualization of student interaction networks, discourse units network and words network (Oshima et al., 2012). Social Network Analysis was used to generate visualization of students' contributions within the discourse and to transform the data into a graphical representation, where measurement indicators such as "density" and "centrality" are used to determine the level of interaction between students (Wortham, 1999). The commonly used SNA has been found to be insufficient in examining community knowledge advancement through students' collaboration and interaction network (Oshima & Oshima. 2007). Interactions between words that learners use in their discourse are important but often neglected, as these words tend to reflect the learners' knowledge and ideas during discourse of a topic. During our analysis using SNA, we used a conventional network measure, betweenness centrality (BC) coefficient, to indicate important connections between words within discourse units. KBDeX graphically displays BC values that allows us to identify and differentiate promising ideas. Our interpretation of the BC coefficient can be viewed as ideas' degree of importance to multiple stakeholders within the discourse at different junctures of the discourse. When viewed over the entire discourse, the BC coefficients form a BC trend line (Figure 2), which suggests the ideas within a DU contain a certain degree of promise towards discourse, and has varying levels of significance and impact on the community discourse after the DU was introduced into the discourse space.
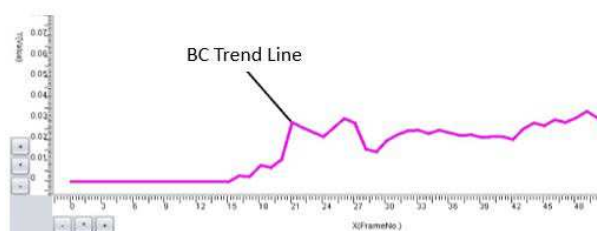
Figure 2. Example of a betweenness centrality (BC) trend line over the entire discourse.

Idea representation through keywords

Ideas are represented mainly through notes on Knowledge Forum, as students seek to query, inform or explain their ideas with their choice of known words. Keywords that are important for learning subject matter, managing learning, and gaining deeper understanding are known as epistemic words. We decided to focus on using objects and nouns to represent tangible items which students could relate easily to, given that the topic on human transport system was pegged at an introductory level. The process of choosing keywords was manually conducted by two experts, who were engaged to determine the list of keywords that reflect the level of knowledge, and are representative of key ideas regarding the topic of human transport system. The final list of keywords were identified with an inter-rater agreement of 82.4%, with differences between the experts resolved after further discussions. The final list of keywords contain 18 unique keywords with their respective 13 plural forms constituting a total of 31 keywords. The keywords are as follows: *red blood cells(s), white blood cell(s), platelet(s), oxygen, carbon dioxide, heart(s), lung(s), blood, blood vessel(s), artery/arteries, vein(s), capillary/capillaries, heart attack(s), heart disease(s), clot(s), antibody/antibodies, oxygenated, deoxygenated.*
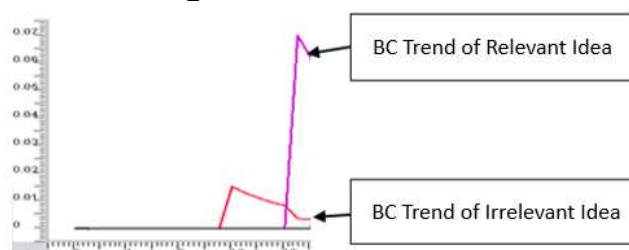
Identifying relevant ideas using I²A

Figure 3. Example of BC trends for relevant and irrelevant idea.

The process of identifying ideas is dependent on the relevancy of contents within the discourse unit. The content of notes in Knowledge Forum form the basis of ideas within the discourse, and the usage of BC trends as a measurement index allows us to compare the relevancy of a discourse unit against other parts of the discourse. High betweenness centrality of a certain discourse unit at a certain juncture of discourse suggests that the selected discourse unit works as a key mediator that links other parts of the discourse through itself, and the idea contributed by the participant might represent an emergent idea that is worth exploring. When there are high values of BC with abnormal behavior within the BC plot, it suggests that a relevant idea could be present and is central at the particular point of the discourse, which learners should pay attention to. On the other hand, discourse units that possess smaller or insignificant values of BC could also represent the existence of less significant ideas that might not be relevant, since there is a lower extent or lack of community talk and discussion regarding the idea and the contents, hence indicating the presence of irrelevant ideas. However, if an idea has promising potential but does not possess sufficient relevancy to the community at the point in discourse, validation through qualitative analysis would then be required to prove and validate the idea's promisingness.

Detecting sustainability of interest in ideas using I²A

As ideas surface and are identified during discourse, it is likely one of two different scenarios could occur to it. If the idea comes across as a major eureka moment of novelty that makes sense and is also interesting to the community, continued discussion among students would most likely occur, with frequent queries and interests to keep the discussion alive with new perspectives, so as to continuously improve the idea and enhance community knowledge. As a result, after the initial spike in BC values, the continuous discovery of new ideas relevant to the original idea will cause the BC trend to decrease gradually with subsequent large BC fluctuations and increments across the remaining parts of the discourse as shown in Figure 4a. The other scenario depicts ideas that fall by the

wayside due to a lack of discussion and community disinterest. These ideas could either be deemed as common topics that were already thoroughly discussed, with convergence of opinions and consensus reached within the community; or it could also mean that potential promising ideas gained less traction than expected within the community. This can be similarly further verified using qualitative analysis. This situation where there is little to no community interest in continuing discussion of the idea at a specific discourse juncture, can be reflected in the BC plot (Figure 4b) with decreasing BC trend that tapers off or reduces to insignificant values.

Fig 4a   Fig 4b
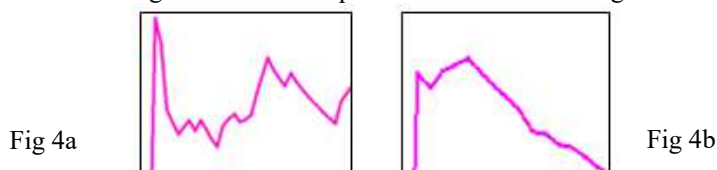
Figures 4a and 4b. Examples of BC trends for promising interesting idea (left) and uninteresting idea (right).

Classification of ideas using I²A

Using the above idea attributes and combining the idea types mentioned in Figures 3, 4a and 4b, our observations in our analysis allow us to classify the pool of ideas within the discourse into four different classes, based on their relevancy and importance to discourse, sustainability of community interest, and impact to the community discourse (Figure 5). Different classes of ideas can surface within a discourse which affects the quality of talk and types of discussion. A relevant idea of sustainable community interest is classified as a promising idea, whereas if the idea has non-sustainable community interest, it is still a potential idea which the teacher can act upon, by engaging students and reviving interest in the topic. However, irrelevant ideas that retain continuous community interests often digress and affect the community's direction in advancing knowledge, therefore requiring teachers to further examine affecting notes and their contents in detail, which can reveal possible distracting elements within the notes. A similarly irrelevant idea with minimal discussions and non-sustainable community interests would otherwise be considered a trivial idea. By classifying these ideas, the summation of the counts of different idea classes within the discourse could also provide a quantitative feel of the discourse's quality.

		Relevancy to the Experts/Teachers	
		Relevant	Irrelevant
Interest to Community	Sustainable	Promising Ideas	Digressing Ideas
	Non-sustainable	Potential Ideas	Trivial Ideas

Figure 5. The four quadrants of idea classification

Findings and discussions

We ran the discourse through I²A using KBDeX for visualization, and were able to classify discourse units (DUs) and the respective contents into different idea types. This was done by using I²A to firstly identify ideas, followed by determination of relevant promising ideas, and lastly conducting an analysis of discourse belonging to different idea types for validation purposes. By tagging and classifying different ideas, teachers and students are able to learn more effectively and efficiently, by choosing to either prioritize improvement of relevant ideas to increase depth in knowledge, or to channel resources into broad based learning of all ideas.

Figure 6 displays three different examples of ideas that were identified throughout the discourse using I²A, along with our qualitative interpretations of how the idea was represented within the discourse unit. These ideas originated from notes extracted from the Knowledge Forum discourse and are an indication of different idea classes and types which can arise from a knowledge building discourse. The DUs were listed in chronological order and need not necessarily be in a conversation with the preceding or subsequent DU, since parallel discussions took place simultaneously across the community, with students responding to different threads over two weeks. The x-axis lists all the DUs within the discourse and are considered as turns in KBDeX, so that step-

wise analysis could be done to track the BC values for each DU, and later used for determining the impact of a certain DU or keyword on the community's discourse. The y-axis indicates the normalized BC value throughout the whole discourse, which reflects the importance of the DU as a key mediator among other DUs, and is also representative of the DU's centeredness within the network. The normalized BC values range from 0 to 1, with a higher score indicating higher centeredness of the DU and its greater role within the discourse. Based on the exemplified DUs and their respective classifications, such a semblance in other discourse can be similarly classified using I^2A for most idea types, and allows the process to be reasonably scalable in longitudinal studies.
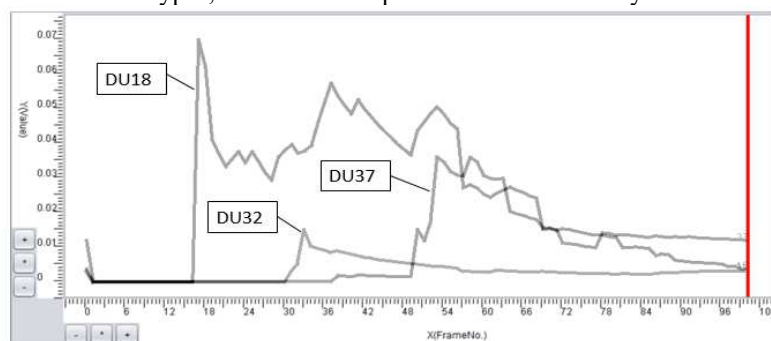

Figure 6. BC graph plot of DU18 (Promising Idea), DU32 (Trivial Idea), DU37 (Potential Idea).

DU32 - Largely irrelevant note with little sustained community interest – Trivial Ideas

S16: my theory - if the flow of oxygen-rich blood to your <u>heart</u> muscle is reduced or blocked, angina (an-ji-nuh or an-juh-nuh) or a <u>heart attack</u> can occur. angina is chest pain or discomfort. it may feel like pressure or squeezing in your chest. the pain also can occur in your shoulders, arms, neck, jaw, or back. angina pain may even feel like indigestion.

Explanation: The idea of DU32 was about physical effects due to blocked passages to the heart, which results in a heart attack. A new term "angina" was introduced, but was never picked up by other students for the remainder of the discourse. Even idea improvements by other students resulted in only a slight increase of BC, but the trend eventually subsided when no other references were made over the rest of the discourse. It was deduced that students reading the contents of DU32 were not spurred to further explore the term "angina" and the underlying reasons for the resulting physical body's reactions. There was little to no interest regarding this idea as DU32 felt more like an informational statement rather than a trigger that stimulates motivation to improve on the idea of angina and its physical effects. Nonetheless, DU32 presented an idea that was slightly interesting to a small portion of the community who have no desire to improve it, and it was verified to be a trivial idea by the teacher.

DU37 - Relevant note with limited interests to community – Potential Ideas

S15: my theory - the components of blood include <u>red blood cells</u>, <u>white blood cells</u>, <u>platelets</u>, and plasma. some blood cells carry <u>oxygen</u> (necessary for metabolic reactions), some blood cells fight off invading substances that could destroy your cells, and other blood cells help to form clots, which keep your body from losing too much blood. the fluid portion of the blood carries nutrients needed to fuel each cell in the body. it also shuttles wastes that need to be transported to the excretory system to be passed out of the body and carbon dioxide that needs to be transported to the lungs to be exhaled. http://www.dummies.com/how-to/content/what-are-the-components-of-human-blood.html ~S15 & S12 –

Explanation: The notes leading up to DU37 provided the community with sufficient materials to craft a general idea (roles of blood components in the human transport system). The idea was improved with extra information about waste movements in DU37 itself and was further substantiated with a web link for verification and exploratory work. Curious students investigated the web link to gather more knowledge about the idea, evidenced by DU51's addition of "blood cell ratios" as new information, and DU54's web link with detailed explanations on "chances of infection". The idea was continuously improved throughout the discourse before a final consensus was reached by the community in the form of a rise above note. The difference between this potential idea and other ideas was the additional work that students put into collating and interpreting of information from previous notes to form connections between keywords. Readers were able to easily link knowledge acquired from other DUs within the current context, and be motivated to continue seeking answers to queries that they generate from reading other students' work. The web link's presence served as an added incentive to search for more evidence

to verify theory proposed by other students. DU37 was considered a culmination of work from rise-above between two students that can provide a more relevant and idea-centric discussion as part of potentially promising ideas.

DU18 – Relevant promising note with high community interest – Promising Ideas

S11: clogged arteries can lead to multiple medical conditions. 1. coronary artery disease. when plaque accumulates in the arteries carrying blood to the heart, it results in coronary artery disease, which could lead to heart attacks. 2. carotid artery disease. the carotid arteries run up either side of your neck and supply oxygen to your brain. when arterial plaque accumulates in carotid arteries, it can lead to stroke.

Explanation: DU18 played an important role in the early part of the discourse as it was the key mediator in linking up two different groups of students that were having separate discussions, but talking about similar ideas such as the definition of arteries and atherosclerosis, but there was a missing aspect of the human transport system that was not discussed, i.e., the blood as a medium that transports resources to other parts of the body and the consequences due to deprivation of such resources. DU18 acted as a trigger that invoked student interests and allowed convergence of ideas into a single note, where students can pool their resources into improving the promising idea. Other than containing part of the idea regarding the circulatory transport system, the causes and effects were also linked together within the same note to prove a causal relationship that made sense to the students. As a result, subsequent discussions were used to describe and explain observations from a different viewpoint, where students were able to start looking at the big picture and strived to improve on the relevant idea instead of focusing on insignificant details. The subsequent contributions were mostly motivated inputs from students that started thinking deeply about the roles of blood constituents and how it affects the human body. The previously mentioned relevant idea in DU37 was directly influenced and traced back to DU18, where the idea had since improved and evolved to become the main idea in DU37.

This study of applying analytics showcases the ability to identify promising ideas currently present within the discourse and early identification of ideas which could evolve into promising ideas in later stages of the discourse. It provides an explanation of an idea's impact on subsequent discourse, since an earlier detected promising idea could have potential influence and ramifications on other ideas later in the discourse. A common observation of the discourse analysis shows that most of the BC values eventually taper off to lower BC values at the end of the discourse, as more notes were introduced into the discourse, and students within the community gradually reached a consensus on key ideas of the topic, resulting in lesser debates or conflicting perspectives regarding ideas, especially in situations where proposed ideas are substantiated with authoritarian sources such as teachers or verified with multiple web resources. To measure the accuracy of automatic tagging of promising ideas, the identified ideas using I^2A were compared against the teacher's list of ideas and keywords that should be acquired by students after the two week learning session. Most identified relevant ideas (88.2%), whether promising or not, were in concordance with the teacher's intention and pre-lesson learning objectives.

Conclusions and future directions

This current study introduces an initial effort in identifying and analyzing promising ideas within the knowledge building environment. By using SNA and measurement indicators such as betweenness centrality, we show that shifting analysis away from the student social interaction network to the idea network of discourse units allows us to identify and assess ideas along with their degree of promise. This work demonstrates the potential for automated tagging of promising ideas and the measurement of subsequent discourse impact on the community, which could significantly affect the community's rise-above. We believe that this research allows students and teachers to focus more on improving identified promising ideas beyond the current status quo, and also provides communities with the chance to advance knowledge by learning more efficiently, rather than spending time trying to identify possible relevant and promising ideas from a large pool of community ideas.

We acknowledge that there would be challenging work in the future, as we seek to expand our studies to students of different levels of expertise such as tertiary level students. We could track the level of understanding and types of ideas within students so as to address the evolvement of ideas and learning processes from the learning journey of a novice into an expert, together with the exploration of possible uncovered idea types. The process of choosing keywords can be enhanced in subsequent studies by including verbs to the list of nouns, as verbs could represent actions and processes that link nouns together, similar to how mind maps function. Another potential direction of research is to further analyze the temporal dimension of the discourse data known as turns in KBDeX, which was briefly mentioned in the paper, so we can understand how SNA measures can affect the discourse with timely interventions. Having analyzed a knowledge building class in practice, future studies can also be conducted not limited to knowledge building communities, but also on other online discourse platforms.

References

Bereiter, C. (2002). *Education and mind in the knowledge age.* Hillsdale, NJ: Lawrence Erlbaum.

Bereiter, C., & Scardamalia, M. (1993). *Surpassing ourselves: an inquiry into the nature and implications of expertise.* Chicago: Open Court.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84.

Chen, B., Chuy, M., Resendes, M., & Scardamalia, M. (2010). "Big Ideas Tool" as a new feature of knowledge forum. *In 2010 Knowledge Building Summer Institute (KBSI).* Toronto, Ontario, Canada.

Chen, B., Scardamalia, M. & Resendes, M. (2013). *Dynamics of Promisingness Judgments in Knowledge Building Work of 8-to 10-Years-Olds.* Poster presented at the 2013 AERA Annual Meeting, San Francisco, CA.

Engeström, Y. (1999). Activity theory and individual and social transformation. In Y. Engeström, R. Miettinen & R.-L. Punamäki (Eds.). *Perspectives on activity theory* (pp. 19-38)*.* Cambridge: Cambridge University Press.

Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine learning, 42*(1-2), 177-196

Lakatos, I. (1970).The methodology of scientific research programmes. In I. Lakatos and A. Musgrave (Eds.), *Criticism and the Growth of Knowledge* (pp. 91-195). Cambridge University Press, Cambridge.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press.

Oshima, J., Oshima, R., & Knowledge Forum Japan Research Group. (2007). Complex network theory approach to the assessment on collective knowledge advancement through scientific discourse in CSCL. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), Proceedings of CSCL 2007 (pp. 563-565). Mahwah, NJ: Lawrence Erlbaum.

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: a social network analysis application for knowledge building discourse. *Educational technology research and development, 60*(5), 903-921.

Paavola, S., Lipponen, L., & Hakkarainen, K. (2004). Models of innovative knowledge communities and three metaphors of learning. *Review of educational research*, *74*(4), 557-576.

Scardamalia, M. (2002). Collective Cognitive Responsibility for the Advancement of Knowledge. In B. Smith (Eds.), *Liberal Education in a Knowledge Society* (pp. 67-98). Chicago: Open Court.

Scardamalia, M., Bereiter, C. (2003). Knowledge building. *Encyclopedia of Education. New York: Macmillan*, 1370-1373.

Scardamalia, M (2004). CSILE/Knowledge Forum. In *Education and technology: An encyclopedia* (pp. 183-192). Santa Barbara: ABC-CLIO.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Eds.), *Cambridge Handbook of the Learning Sciences* (pp. 97-118). New York, NY: Cambridge University Press.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher, 27*(2), 4-13.

Strijbos, J. W., Martens, R. L., & Jochems, W. M. (2004). Designing for interaction: Six steps to designing computer-supported group-based learning. *Computers & Education*, *42*(4), 403-424.

Sun, W., Zhang, J., Jin, H., Lyu, S. (2014). Analyzing Online Knowledge-Building Discourse Using Probabilistic Topic Models. *International Conference of Learning Sciences*, 2, 823-830.

van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning*, *4*(3), 259-287.

Wortham, D.W. (1999). Nodal and Matrix Analyses of Communication Patterns in Small Groups. In C. M. Hoadley, & J. Roschelle (Eds.), *Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference* (pp.681-686)*.* Palo Alto, CA: Stanford University.

Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge building communities. *Journal of the Learning Sciences, 18*(1), 7–44.

Acknowledgments