DIME: A Dynamic Interactive Mathematical Expression Tool for STEM Education

Mr. Donald Joseph Beyette, Texas A&M University

Donald Beyette is a master thesis student at Texas A&M University studying machine learning, graph theory, and GPS navigation. Current research projects focus on content analysis, systems to model users learning behavior, hypersonic navigation, and GPS antispoofing techniques.

Mr. Michael S. Rugh, Texas A&M University

Michael S Rugh is a second year PhD student focusing on mathematics education within the Curriculum and Instruction PhD track in the Department of Teaching, Learning, and Culture within the College of Education and Human Development at Texas A&M University.

Mr. Jason Lin, Texas A&M University

Jason Lin received his Master's degree in Computer Science from the National Cheng-Kung University, Tainan, Taiwan, in 2011. He is currently pursuing a Ph.D. degree with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX, USA. His research interests include Scientific Document Analysis, Geospatial Intelligence, and Quantum Cryptography.

Mr. Xing Wang, Texas A&M University

Xing Wang is a Ph.D. student in the Department of Computer Science and Engineering. He pursues research in math- centered publication analysis at the Real-Time Distributed System lab supervised by Dr. Jyh-Charn (Steve) Liu. Xing received his Bachelor's and Master's degrees in computer science from Peking University, China.

Mr. Zelun Wang, Texas A&M University

Zelun Wang received his Bachelor of Engineering degree in the Department of Automation from Xi'an Jiaotong University, Xi'an, China, in 2014. He is currently working towards the Ph.D. degree in computer science with Dr. Jyh-Charn Liu at Texas A&M University, College Station, TX, USA. His research interests include machine learning and scientific document analysis.

Dr. Jyh-Charn Liu, Texas A&M University

Dr. Liu earned his PhD from University of Michigan in 1989. Since then he has worked in the department of Computer Science and Engineering, Texas A&M University. His primary research interests span from computer systems, system modeling, and to the future generation of real-time, distributed systems. He is keen on developing innovative technologies to enable students in overcoming the common learning barriers, especially the STEM fields.

Dr. Robert M. Capraro, Aggie STEM @ Texas A&M University

Robert M. Capraro, is Co-Director of Aggie STEM, Director of STEM Collaborative for Teacher Professional Learning, and Professor Mathematics Education in the Department of Teaching Learning and Culture at Texas A&M University. Dr. Capraro's expertise is applied research in school settings, program evaluation, the teacher as change agent for STEM school improvement, and STEM student achievement. He recently received the best paper award from the International Conference on Engineering Education where he and two colleagues presented their work related to the Aggie STEM project. He is currently involved in research in four school districts and more than 20,000 students and 80 teachers. His editorial work includes Associate Editor of the American Educational Research Journal, School Science and Mathematics, and Middle Grades Research Journal and the Research Advisory Committee for the Association of Middle Level Education. He was selected as a minority scholar for 2007 by the Educational Testing Service and served as president of the Southwest Educational Research Association. He is the author or

co-author of three books, several book chapters and more than 100 articles on mathematics education, quantitative research methods, and teacher education published in such venues as Journal of Mathematics Education, International Journal for Studies in Mathematics Education, Journal of STEM Education: Innovations and Research, International Journal of University Teaching and Faculty Development, LEARNing Landscapes, Special Issue: Mind, Brain and Education, Journal of Mathematical Behavior, European Journal of Psychology of Education, The Journal of Mathematical Sciences and Mathematics Education, Urban Review, Journal of Urban Mathematics Education, Educational Researcher, Cognition and Instruction, Educational and Psychological Measurement. He recently was awarded a $400,000 dollar grant - - continued support by the Texas Higher Education Coordinating Board to continue his work with developmental education bringing his total external funding in excess of 15 million.

DIME Map: A Dynamic Interactive Mathematical Expression Tool for STEM Education

Abstract

The learning of abstract relationships in Science, Technology, Engineering, and Mathematics (STEM) fields for students is an issue that is complex and riddled with difficulties. Current educational e-books and graphic organization systems attempt to illustrate those difficult concepts, but are limited due to the single reasoning perspective created by the author(s). Thus, we discuss how the computer generated Dynamic Interactive Mathematical Expression (DIME) system serves as an adaptive educational technology for students and teachers. This system creates a dynamic inter-dependent graphic visualization of interrelated mathematics topics (the DIME map) from PDF files by automatically extracting mathematical expressions and declarations from the material. Students and teachers can rearrange the DIME map to provide their own unique visual representation of the mathematical expressions. Each student provides feedback that allows the DIME map to adapt to their learning patterns and to remediate potential struggles. Through this paper, we present a new framework for the DIME system, which adapts to individual teaching and learning situations by integrating feedback from teachers and students into the DIME maps feature network.

Introduction

Graphical displays of information have been used to communicate, educate, and maintain cultural records of important information throughout human history. Changing with time and technology, graphical displays of information have evolved from cave paintings [8], Ancient Egyptian art (3000BC-30AD) [9], papermaking in China (25-220 CE) [10], and paper books to modern digital documents and videos. In 1968, David Ausubel proposed that the use of graphic displays of information (e.g. flow charts, semantic diagrams, and tree maps) would greatly assist students in understanding new, interrelated, and potentially complicated information [24]. By June 2018, digital documents, videos, and computers were available for about 55.1% of the people in the world [11]. Thus, graphical displays of information have become more relevant in education, because of the increased availability of computers worldwide. Studies have shown that online sources of information, such as textbooks, increase the average performance of students for quizzes by about 16% [5]. Interactive e-books show a 16% increase in exam grades, 7.4% increase in project grades and 16% more lesson effectiveness than students who used traditional textbooks [6]. Academic organizations [15] [16] [17] and industry [18] [19] [20] [21] [22] [23] continue to invest into developing more effective and diverse online interactive textbooks, as studies continue to show large positive differences between traditional and online interactive textbooks. Graphical tools and e-books for different STEM subjects aim to improve students' interest and ability to understand complicated concepts and connections [1] [3] [4] [7]. Some online interactive textbooks record each student's activity for reading assignments and quizzes, but does not integrate this information into the system [7] [23]. These graphical tools and e-books focus on a single subject with multiple modules explaining each case, requiring vast amount of effort from an expert to create such modules. As a result, the static information is unable to operate optimally in new environments, because the system is unable to adapt uniquely to each students' learning behaviors and teachers' input. Thus, to overcome this issue, the DIME system integrates modern machine learning techniques, which allow the graphical display of information to adapt automatically based on external and internal feedback.

The way the DIME system responds to these two types of feedback differentiates it from previous systems. Internal feedback refers to the students' physical interaction with the DIME map, such as clicking, dragging, and searching patterns. These internal interactions alone are not enough information to modify the DIME map optimally, as it is not possible to separate actions based on curiosity and those purposely based on learning objectives. Thus, external feedback such as the course learning objectives, test scores, and assignment scores will ensure that the map optimally adapts to the specific learning situation. However, if external feedback is not available, the user will be encouraged to inform the system what their current learning objective is (e.g. identify key sections, terms, and mathematical expressions). A feature network will be used to combine feedback with the DIME map to dynamically modify the graphical display of information. As a result, the DIME system is able to work with any subject that contains mathematical expressions (such as variables and equations). The remaining sections will briefly describe the DIME system, previous findings regarding the system, and the feature network that is currently being integrated into future iterations of the system.

The DIME System

The DIME system is a web-based application that uses MECA (Mathematical Expression Content Analysis) software to construct a DIME map that contains extracted mathematical expressions, semantics, and declarations from a PDF file [12] [13] [14].

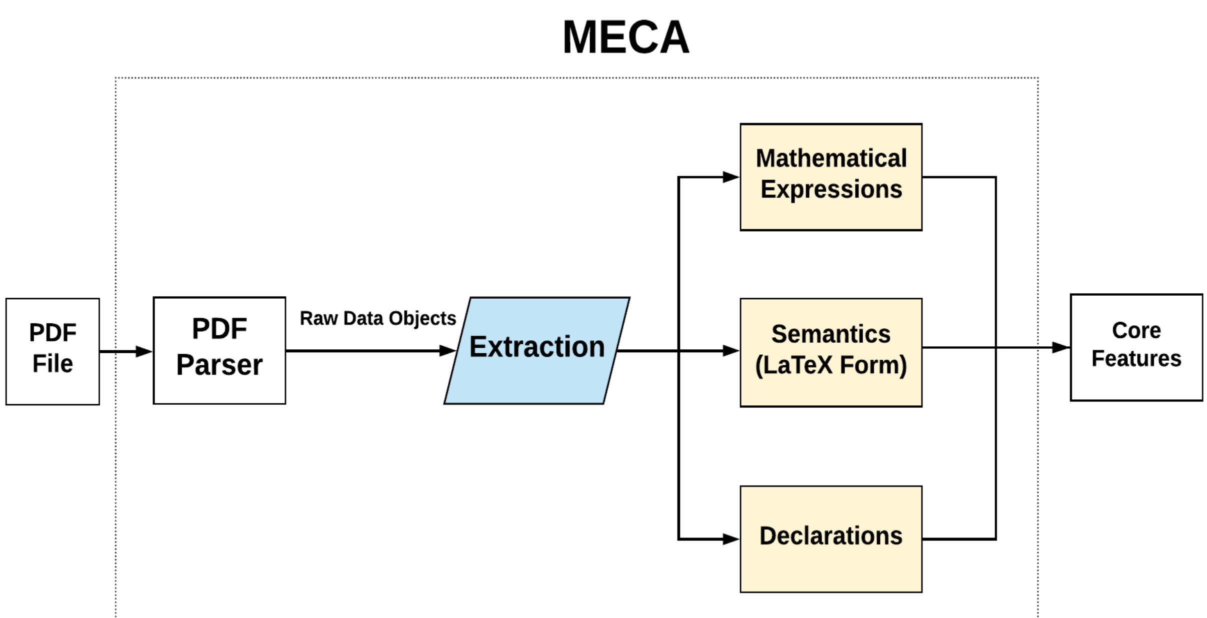

Figure 1: Extraction of Core Features from MECA

Mathematical expressions are a collection of characters from the PDF file that represent expressions such as equations and variables (e.g. $\omega, \frac{1}{2}at^2$, and $\overrightarrow{d_j} = p_{j0} \cdot \overrightarrow{t_0} + p_{j1} \cdot \overrightarrow{t_1} + \cdots + p_{jn} \cdot \overrightarrow{t_n}$). Alone, these collections of characters cannot provide enough information to build relations between mathematical expressions, because their original syntax and meaning is unknown. For example, the equation $y = b^2 * \beta$ is grouped as [y, \eq, b, *,\Beta, 2], which does not provide enough meaning to correctly recover the equation. Thus, semantics (object meanings, spatial layout, and order of operations) are used to generate the LaTeX structure of each mathematical expression [14]. Lastly, declarations represent word meanings and word relations for each mathematical expression such as a name or definition. Collectively, the mathematical expressions, semantics, and declarations created by MECA define the Core Features used to initialize the DIME feature network shown in Figure 2. Users will upload a PDF file to DIME, which will then communicate with MECA in order to retrieve the Core Features from the document. Afterwards, DIME will remain in a feedback loop, which automatically modifies the DIME map based on internal and external features.

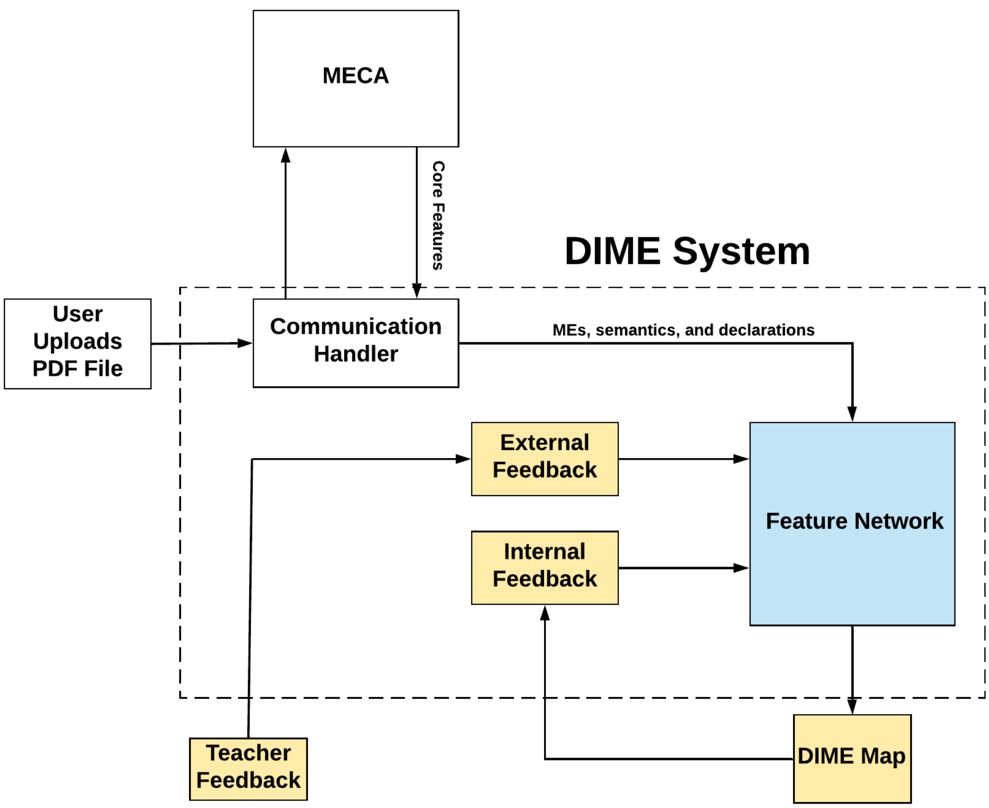

Figure 2: Overview of DIME. The dashed line represents the DIME system that is repsonible for generating and maintaing the DIME Map.

The initial state of the DIME map is determined by the Core Features that are received from the MECA system. These Core Features create a network of nodes and edges, which collectively make up the DIME map. Each node represents a mathematical expression and its declaration, while edges define the relationship between them. Relations between nodes are determined by identifying whether or not a particular mathematical expression is on the left or right hand side of an equation. For example, the system identifies the equation $F = -k \cdot x$, as well as the individual variables, k and x, which are introduced in other places in the document. Because the variables, k and x, are on the right side of the equation, the system interprets them as building into the equation, and therefore into the variable F. In the resulting map, arrows from both k and x to the equation would demonstrate these directional relationships shown in Figure 3.

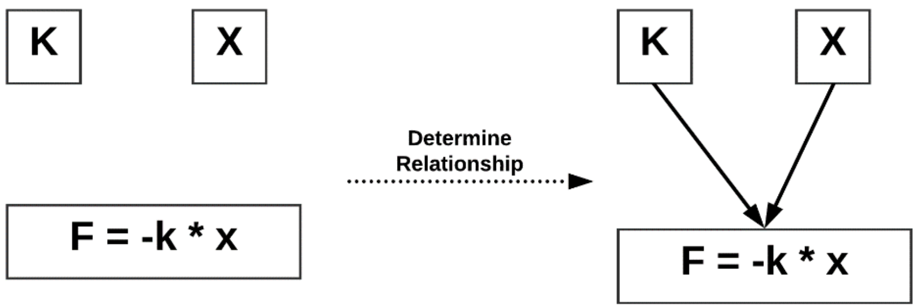

Figure 3: Building relations between nodes

The finished DIME map will contain a vast network of these nodes and edges, allowing the student to follow the flow of information in the book easily. Figure 4 shows an example DIME map (on the right) which is displayed next to the e-book version of the uploaded PDF file. The numbers in Figure 4 correspond to the internal IDs in Table 1, which provide internal interactive feedback to the DIME network.

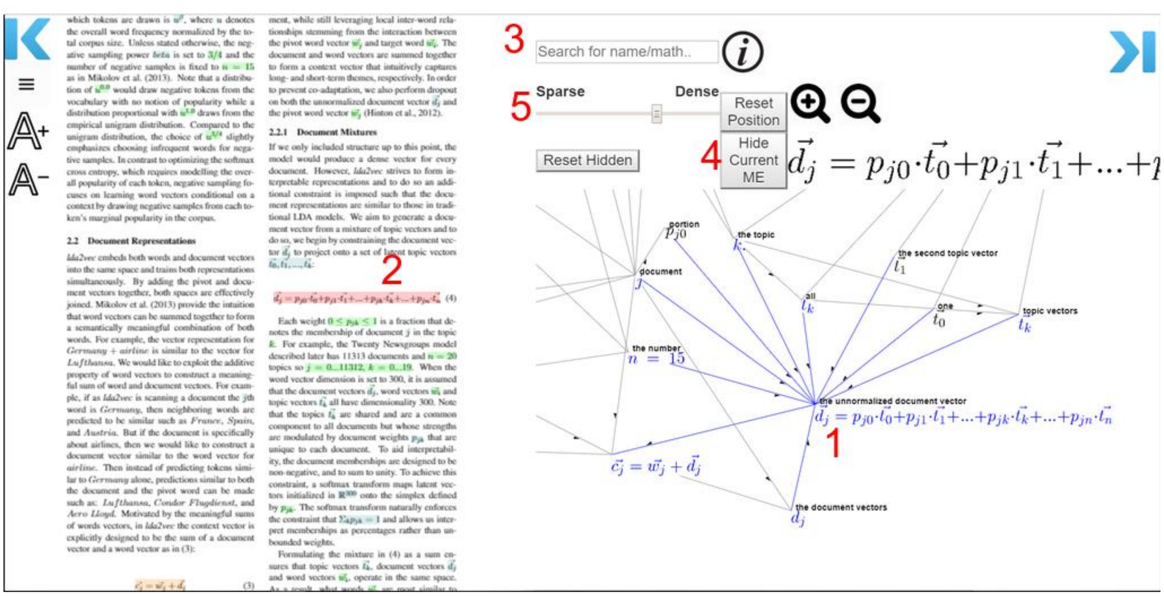

Figure 4: The numbers in this figure refer to internal features from Table 1.

Some additional features include highlighting all direct edges (relationships) when the user hovers over a specific node (mathematical expression), seen in Figure 4.

Table 1: Key Internal Features

ID	Type	Description
1	Internal	Custom arrangement of map nodes by clicking, holding, and dragging the left mouse button on an individual node.
2	Internal	Navigation between an object in the map and its location in the book by clicking on the node.
3	Internal	Search feature that uses declarations, LaTeX language, and symbol names.
4	Internal	Hide nodes that the user is not interested in.
5	Internal	Increase or decrease the distance between nodes.

Together, these features allow the student to visualize and interact with the connections between interrelated mathematical expressions. Students are able to follow the teacher as they explain the bigger picture by visualizing what each mathematical expression builds to. The graph can act as a navigator for the text book, which provides information about mathematical expressions, definitions, and how it relates to other concepts not learned yet.

Experimental Pilot Study

A previous iteration of the DIME system was tested in a small pilot study [15]. A quasi-experimental, pretest-posttest, control group design was used with high school students attending a STEM summer camp. The experimental group used the DIME map to investigate the relationships between physics topics involved in fixed axis rotation. Some students experienced initial problems with using the system due to computer incompatibility, the DIME system latency, and complexity of appearance. When these problems occurred, the students often worked together to explore and manipulate the DIME map on a single device. These issues have been addressed and improvements are continuously being made. Therefore, we expect the number of students independently using the DIME map to increase greatly in future implementations, as future iterations will aim to solve the problems experienced in this quasi-experimental study.

This pilot study found promising results on the effects of high school students using the DIME map to learn advanced physics concepts [15]. The results of that study also showed that the experimental group scored higher than the control group on their ability to make connections between interrelated mathematical expressions and on their attitude toward learning science [15]. Females experienced over double the effect that males did in these outcomes from using the DIME map. These findings demonstrate the potential of the DIME map as an equitable and empowering educational tool for students and teachers. The feature network presented in this paper is based on the results of this small pilot study and is used in current iterations of the DIME map.

Feature Network

The feature network controls the orientation and physical attributes of every node and edge in the DIME Map. Physical attributes such as size, color, and opaque level target each node (mathematical expression and declaration), while the orientation changes the nodes position. Core, internal, and external feature weights will determine both the orientation and attributes of every node in the map. Core weights are determined during the extraction and analysis stage of MECA, which defines the initial feature network. Each core feature list in Table 2 will be assigned initial weights based on MECA metrics [12] [13].

Table 2: Core Features

ID	Type	Description
1	Core	Extracted **mathematical expressions** from PDF file
2	Core	Extracted **semantics** from PDF file
3	Core	Extracted **declarations** from PDF file

Weights are used to modify the DIME map by directly influencing each nodes orientation, size, color, and opacity describe in Table 3.

Table 3: Node Attributes

Type	Description
Orientation	Position of the node with respect to every other node in the map, which is determined by a node’s weight and feature score.
Size	Increases as the weight of the node increases.
Color	Changes based on searching, mouse interactions, and external feature weights.
Opacity	Opaque scale from 0-1, where 0 makes the node hidden.

Out of the four attributes listed in Table 3, orientation is the most critical attribute of the DIME map, as this controls the flow between concepts. As mentioned in the orientation description, weight is used to determine the position of the node. This is the result of the force between each node, which works similar to the relationship between planets and gravity. Larger nodes (like planets) will pull nodes closer to them, especially nodes that share a strong relationship with each other (same domain). Internal and external feature weights will directly affect the weight of each node, causing nearby nodes to periodically change their orientation and size as the student progress in the class. The system will notify the student after enough feedback is received to modify the DIME map, and the student can choose to accept or reject the modification.

Internal and external features modify the weight of each nodes Core Features, causing the map's orientation to change. All internal features receive weights based on the total distribution of interactions with internal features 1-5 from the first experimental group listed in Figure 5. Each internal feature weight is calculated from equation (1), where the superscript *I* represents internal features and *n* the feature ID.

$$W_n^I = \frac{\text{Total number of user interactions with internal feature } n}{\text{Sum of all user interactions on every internal feature}} \quad (1)$$

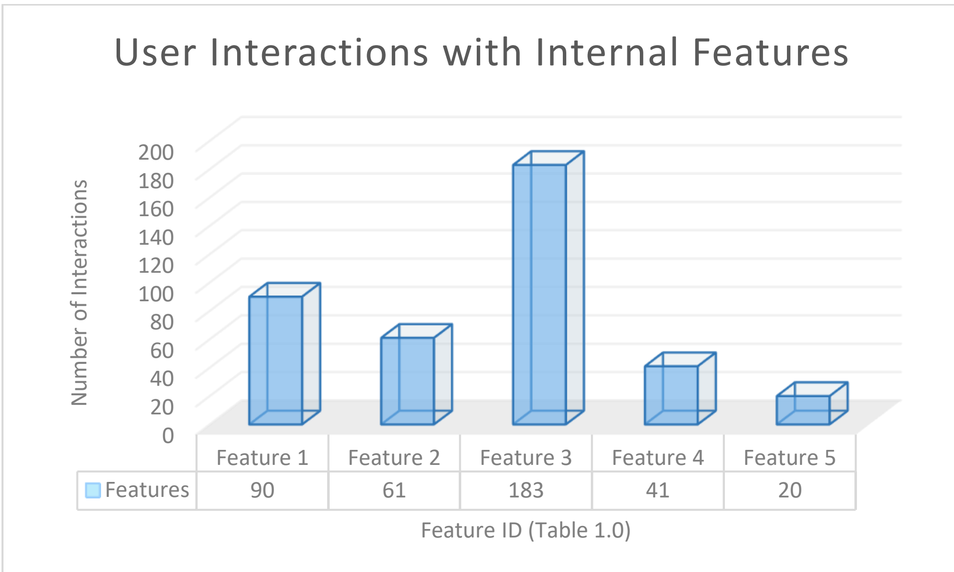

Figure 5: Refer to Figure 2 for each feature ID (1-5) and Table 1 for the description.

From Figure 5, searching will receive more weight than the other features, as students used this more often. This result is logically correct, as it implies students must be searching with specific words related to their learning objective rather than playing with the graph. We expect the distribution of weights for each feature to stabilize as larger experimental groups are used. Future experiments may show different weight values for each feature based on the participant or setting characteristics. For example, high school and college students may prefer different features such as, navigating between the e-book and map instead of using the search bar. Students will receive an internal node score for each mathematical expression based on the weights from (1), described below in equation (2). Node scores determine the nodes overall weight, which is then compared to every other node.

$$EF_S^I = \sum_{i=1}^{n} N_i * W_i^I \quad (2)$$

EF_s^I represents the internal nodes score for student s, N_i represents the total number of interactions for feature i, and W_i^I defines the weight for internal feature i. Every mathematical expression will receive a EF_s^I score, as this is required to update the map by modifying the feature network (Refer to Figure 2). Weights from internal features may create bias, as students could be exploring or playing with the map rather than spending more time with a particular feature due to focus or study. Thus, external features are used to enhance and dampen the effects of internal features. Table 4 describes external features that are determined by the teachers' input about each individual user's interaction or performance in the class. For example, if the class curriculum is focused on acceleration, then students who struggle with this concept on quizzes, homework's, and exams will have increased weights for mathematical expressions related to acceleration. This means that concepts in future sections or chapters that relate to acceleration will receive additional weight. Equation (3) calculates the new core feature weight (CF_i), where i is the ID for some core feature in Table 2 (mathematical expression, semantics, declaration).

$$CF_i = CF_i * W_n^E \qquad (3)$$

The external feature weights' (W_n^E) impact on the DIME system compared to the internal features impact is greater, as the external features are based on direct teacher feedback.

Table 3: External Features

ID	Weight	Type	Description
1	W_1^E	External	Scores from test, home works, and quizzes as $\frac{score}{total\ points}$ on a scale of 0-1.
2	W_2^E	External	Direct feedback from the teacher's observation on a scale of 0-1, with 1 being very easy and 0 very hard.
3	W_3^E	External	Importance on a scale of 0-1, with 0 being irrelevant and 1 being extremely important. (How strong does the material relate to the current learning objective?)

Each external weight is calculated by subtracting the score (e.g. importance, test, or quizzes) provided by the teacher from one, shown in equation (4). S is on an interval from [0-1].

$$W_n^E = 1 - S \qquad (4)$$

Students who receive high scores will have lower external weights, as a perfect score of one will result in an external weight of zero ($W_n^E = 1 - 1 = 0$). Thus, the core feature weight for a specific mathematical expression will decrease as students perform better in class and on exams.

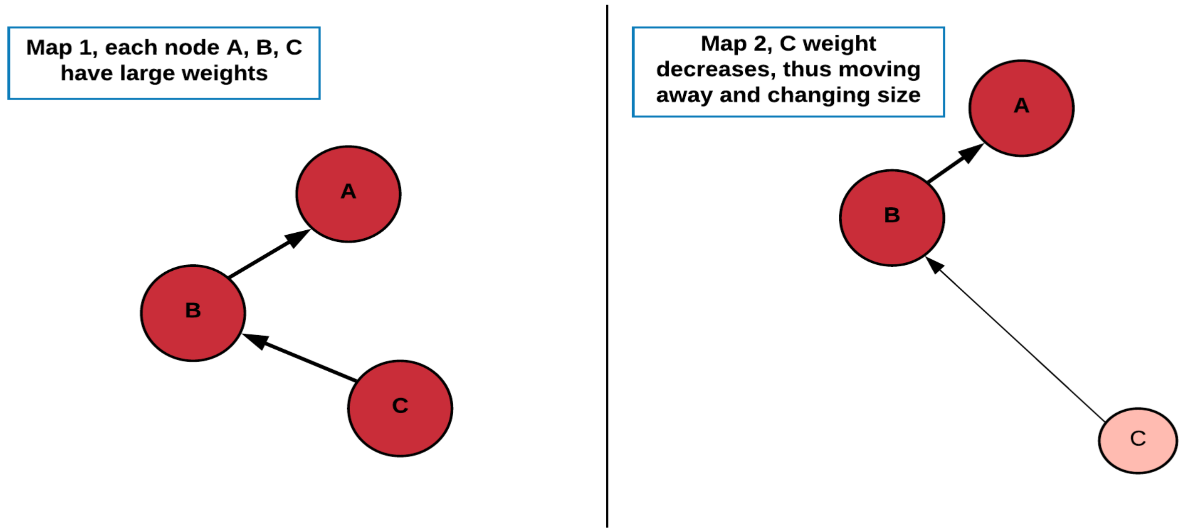

Figure 6: Map weight change example

Figure 6 describes a simplified DIME map, where A, B, and C represent mathematical expressions where the student is currently struggling. As the student performs better on mathematical expression C, the nodes weight decreased, causing it to move away from the more difficult mathematical expressions where the student is struggling. The DIME map uses the core, internal and external feature weights to arrange these nodes into a graphical display (Figure 4). The resulting map is a network of nodes and edges based on the students' learning behaviors and interactions with the system. Students and teachers may still change this map into a representation that best fits them by using the features described in Figure 4 and Table 1 (such as dragging and hiding). Figure 7 represents a manual modification of the DIME map by a teacher from the pilot study. The teacher was focused on teaching rotational kinetic energy and angular momentum, which is located at the bottom of the map. While the ability to rearrange nodes is unique to this technology, the teacher's arrangement of the nodes is reminiscent of traditional, teacher-made concept maps. The purpose of introducing feedback and weights is for the system to automatically create something meaningful that inherently helps guide the student's learning process. Future iterations of the DIME system will strive to automatically generate a similar map to those created by the teacher by using external feedback to prioritize the targeted rotational kinetic energy and angular momentum nodes. When the teacher specifies which nodes are most important, the system increases the targeted node's weight and the weight of any nodes that share a connection with the target node. The system will then traverse through every node, analyzing the strength of its connection to the targeted nodes and their variables. Collectively, the DIME map will utilize a feedback loop that adapts based on the integration of equations (1), (2), and (3) with the base feature network. The full extent of the feature network is not discussed in detail because the write-up would be extensive and out of the scope of this paper. The fundamentals discussed in this paper demonstrate the power of the DIME system to automatically create a map of knowledge and display it in a meaningful, logical way by utilizing the feature network.

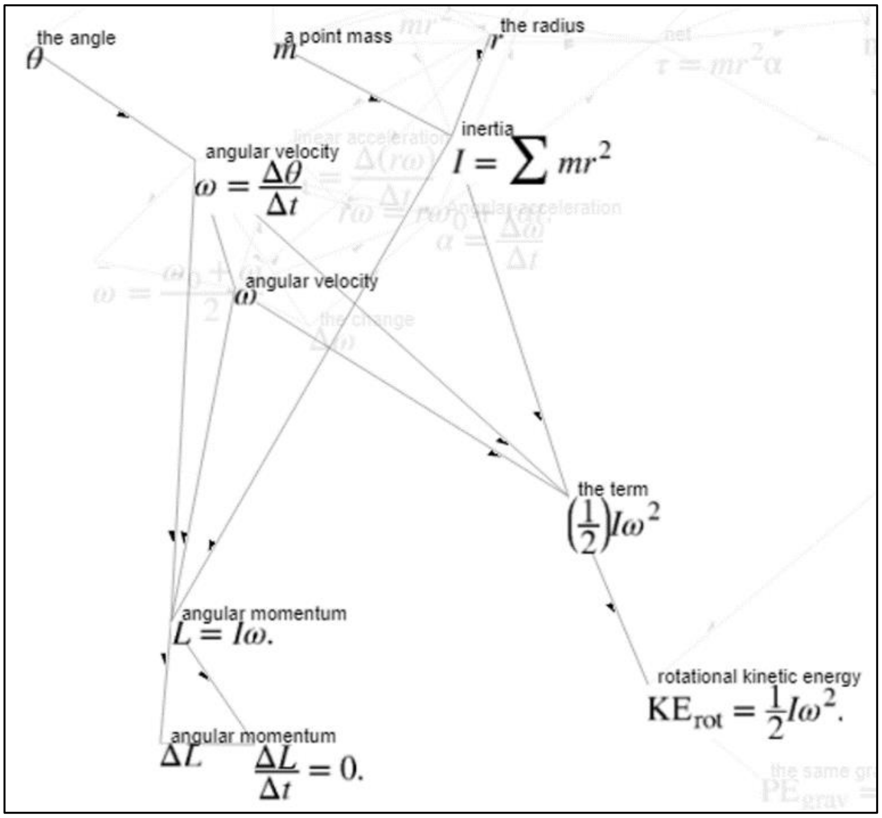

Figure 7: A teacher who was preparing for a presentation modified this DIME map with the dragging feature (Internal feature 1) and the hide feature (Internal feature 4).

Conclusion

The DIME map has great potential as a tool for students to visualize and understand connections between interrelated mathematical expressions. By overcoming the complexity of appearance issue with students in the pilot study, future iterations of the DIME system will be able to provide a unique way of approaching the structure of knowledge for students through a feature network that is modeled based on their unique interactions with the DIME map. We are currently testing the latest version of the DIME system on a larger sample size of students and longer time duration to obtain a robust estimate of effectiveness. Future iterations of the system will provide the teacher with suggestions on what the student may be currently struggling with or what they might struggle with in the near future. In addition, the DIME map will develop unique features for each student based on their feature network, which is determined by that student's interaction with the DIME map and performance in the course (teacher feedback). Students and teachers can use the DIME map with any STEM material that contains mathematical expressions, making it a diverse tool for multiple subjects.

Reference List

[1] Ngo, C. C., and K. Gramoll. "A Web-based Electronic Book (eBook) for Fluid Mechanics." age 9 (2004): 1.

[2] Learning with Python: Interactive Edition 2.0. http://interactivepython.org/courselib/static/thinkcspy/index.html. February 2015.

[3] Huang, Meirong, and Kurt Gramoll. "Online interactive multimedia for engineering thermodynamics." *In ASEE Annual Conf. Proc., Salt Lake City, UT*, pp. 20-23. 2004.

[4] Gramoll, Kurt. "A web-based electronic book (ebook) for solid mechanics." *In Proc. ASEE Annual Conf.(Mech. Div. Best Pres. Award), Honolulu, HI, USA. 2007*.

[5] Edgcomb, Alex, and Frank Vahid. "Effectiveness of online textbooks vs. interactive web-native content." *In 2014 ASEE Annual Conference*. 2014.

[6] Edgcomb, Alex, Frank Vahid, Roman Lysecky, Andre Knoesen, Rajeevan Amirtharajah, and Mary Lou Dorf. "Student performance improvement using interactive textbooks: A three-university cross-semester analysis." *In 2015 ASEE Annual Conference and Exposition.* 2015.

[7] Matthew W Liberatore, and New Orleans Jazzed Engineering Education. "An Interactive Web Native Textbook for Material and Energy Balances". *AEEE's 134rd Annual Conference & Exposition, New orealns, LA. June 26-29, 2016.*

[8] Robert Gunn, Bruno David, Jean-Jacques Delannoy and Margaret Katherine, "The past 500 years of rock art at Nawarla Gabarnmang, central-western Arnhem Land" in: Bruno David, Paul S.C. Taçon, Jean-Jacques Delannoy, Jean-Michel Geneste (eds.), The Archaeology of Rock Art in Western Arnhem Land, Australia (2017), pp. 303–328.

[9] The Art of Ancient Egypt. A resource for educators. New York: The Metropolitan Museum of Art. P. 44. Retrieved July 7, 2013.

[10] Tsien, Tsuen-Hsuin (1985). "Paper and Printing". Joseph Needham, Science and Civilization in China, Chemistry, and Chemical Technology. 5 part 1. Cambridge University Press: 38.

[11] World Internet Users Statistics and 2016 World Population Stats. [Online]. Available: https://www.internetworldstats.com/stats.htm [Accessed Feb. 2th 2019]

[12] X. Wang, J. Lin, R. Vrecenar, and J.-C. Liu. "QuQn Map: Qualitative-Quantitative mapping

Of scientific papers" *In Proceedings of the 18th ACM Symposium on Document Engineering*. Halifax, Nova Scotia, Canada, Aug. 28-31, 2018.

[13] J. Lin, X. Wang, and J.-C. Liu. "Prediction of mathematical expression constraints (ME-Con)" I*n Proceedings of the 18th ACM Symposium on Document Engineering.* Halifax, Nova Scotia, Canada, Aug. 28-32, 2018.

[14] X. Wang, J. Lin, R. Vrecenar, and J.-C. Liu. "Syntactic role identification of mathematical expressions". *Proceedings of the Twelfth International Conference on Digital Information Management*. pp. 179-184, Fukuoka, Japan, Sept 12-14, 2017.

[15] Rugh, M.S., Wang, X., Lin, J., Barroso, L., Capraro, M. M., & Capraro, R. M. (2019, April). Computer generated dynamic graphic organizer's effect on learning outcomes. Presented at the national Council of Teachers of Mathematics Research Conference, San Diego, California.

[16] OpenDSA: Open Source Interactive Data Structures and Algorithms. [Online]. Available: http://algoviz.org/OpenDSA/. February 2015. [Accessed Feb. 2th 2019]

[17] Open Learning Initiative at Carnegie Mellon University. [Online]. Available: http://oli.cmu.edu/. February 2015. [Accessed Feb. 2th 2019]

[18] Boundless. February 2015. [Online]. Available: https://www.boundless.com/. [Accessed Feb. 2th 2019]

[19] Flat World Knowledge. September 2014. [Online]. Available: http://catalog.flatworldknowledge.com/. [Accessed Feb. 2th 2019]

[20] For Dummies eLearning. February 2015 [Online]. Available: https://learn.dummies.com/. [Accessed Feb. 2th 2019]

[21] Inkling. February 2015. [Online]. Available: https://www.inkling.com/. [Accessed Feb. 2th 2019]

[22] WileyPLUS. February 2015. [Online]. Available: https://www.wileyplus.com/. [Accessed Feb. 2th 2019]

[23] zyBooks. February 2015. [Online]. Available: https://zybooks.zyante.com/. [Accessed Feb. 2th 2019]

[24] Ausubel, D.P. (1968). Educational psychology: A cognitive view. New York, NY: Holt, Rinehart, & Winston.