Student-made Models Analyzed on a Large Scale (SMALS)

PI: Caleb Trujillo
School of Interdisciplinary Studies
University of Washington Bothell
eGC1 A173149

Abstract

Classroom learning is typically assessed using multiple-choice or written responses that are scored on a linear scale from 0% to 100% and, despite calls for using scientific practices as indicators of learning, most exams in STEM classrooms focus on knowledge- and fact-based instruction with little attention to alternative assessments that reveal knowledge restructuring in a non-linear way. In this proposal, we seek to use a network science approach in an educational context to analyze student-constructed models of biological systems and statistical concepts (1) to explore changes in models as evidence of learning, and (2) to compare the observed changes in student models to the simulation of network growth dynamics. Using network analysis, learning theories of knowledge structures, and a collection of concept models made by students as assessments in biology and statistics coursework, this project seeks to forward an understanding of concept models as representations for measuring learning in a way that can show development at both the individual and collective scales.

2. Response to Previous Reviews

First application to RRF.

3. Description of Proposed Research

A. Introduction and Rationale

Our project proposes to use a network science approach in an educational context to analyze student-constructed models of biological systems and statistical concepts (1) to explore changes in models as evidence of learning, and (2) to compare the observed changes in student models to the simulation of network growth dynamics. The project prioritizes interdisciplinary work, strengthens educational research across institutions between UW-Bothell and Michigan State University, and seeds an externally-funded research program. This project works across disciplines by using research techniques from data science fields, leveraging theory from cognitive psychology, and working in discipline-specific education for biology and statistics to address major gaps in education research outlined in the following sections. Together, we expect that evidence gleaned from this research will inform an expanding research program to use network approaches for studying discipline-specific teaching and learning and disrupt traditional approaches for assessing learning.

Gap 1: Data science approaches are needed to better assess domain-specific learning with non-linear measures.

Many sectors like business, health informatics, and life science research are using data-intensive approaches to generate insights, but similar advances in education research are uncommon. Therefore, education needs to grow its capacity for data-intensive research by building large datasets, adapting to the expansion of learning technology, and implementing data science approaches to understand fundamental aspects of teaching and learning including differences among learners [1,2]. In undergraduate science education, projects that apply data science approaches to investigate student learning in science classroom settings are rare [3,4] and they often depend on traditional assessment data (i.e. multiple-choice and written response). Furthermore, most educational tests evaluate students on a linear scale from zero to 100 percent and assume that students with a higher score have a deeper understanding of the domain. Such assessment strategies are useful for providing a snapshot of student learning at a given moment within a course or curriculum but overlook two facts about learning:

1. Learning is not linear
2. A “final” state of knowledge about a domain does not exist.

An overreliance on fact- and knowledge-based assessments is in opposition with recommendations from leaders and evidence-based education research that advocate that students engage in meaningful practices as a way to learn science [5–7]. Student-made models provide an alternative, understudied form of assessment that engages students in constructing representations of their learning that are non-linear, show complexity, and enable measurement of the restructuring of knowledge. To address the scholarly gap, our research poses the following question: ***How can network approaches be used to assess learning evident in models made by students in biology and statistics***?

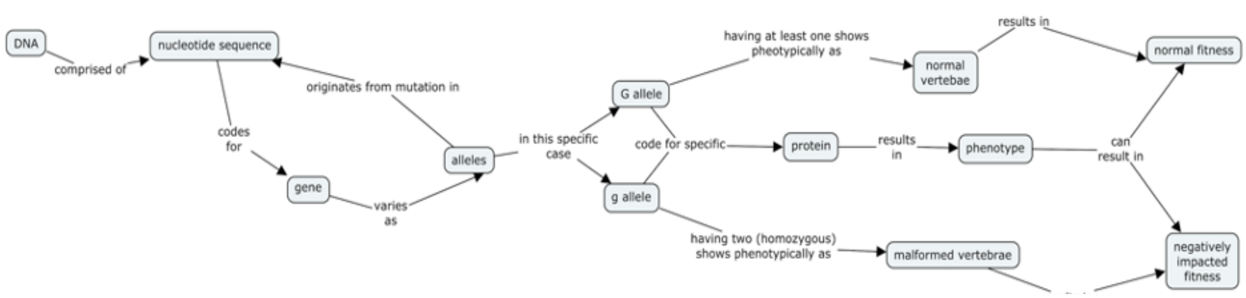
Gap 2: The lack of clarity about network growth dynamics for cognitive models.

Concept maps and student-made models visualize concepts and relationships among concepts and therefore resemble semantic networks consisting of nodes and edges. However, with few exceptions, developed network analysis techniques [8,9] are largely unexplored as a way to understand such educational representations and the structural changes that occur as a student or group of students learns. While few studies have examined the network metrics of student models, even fewer have compared these changes to the simulations of network growth such as small-world structure, hierarchies, and scale-free networks [8]. The use of network growth dynamics to understand student learning of specific domains presents an uncharted research front. This study poses the following question: ***How can simulations of network growth dynamics support an explanation of the learning processes evident in student models?***

B. Objectives

1. Explore the use of network approaches as a means to assess learning evident in models made by students in biology and statistics.
2. Compare the growth evident in student models to simulations of network growth dynamics to explain knowledge restructuring occurring during domain-specific learning.

Student models provide a window into learning.
Over the last 13 years of using models during instruction, the research group at Michigan State University collected over 13,000 models made by undergraduate students in biology (e.g., Figure 1). Now, over 180 models have been collected from statistics courses at UWB as a way to assess students' cognitive connections of important concepts. These models are useful because they provide an easy assessment that can capture nuances of student understanding in a unique representation. This massive set of data surrounding representations provides a unique opportunity to apply data science methods to generate new insights about models as a non-linear assessment of learning, a teaching tool, and a representation of knowledge restructuring.



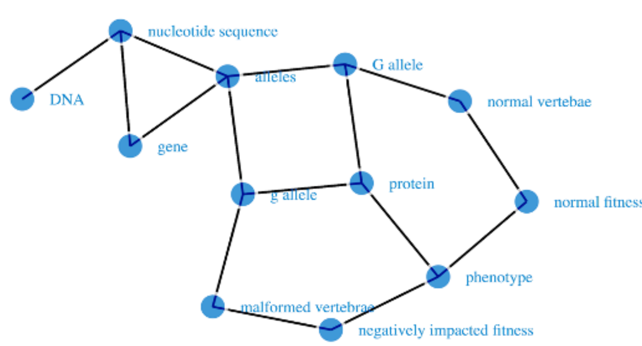


Figure 1 (top): A student's model composed of concepts shown in boxes and relationships shown as arrows between boxes to explain how genetic variation can lead to phenotypic differences in the context of the malformed vertebrae of wolves.
Figure 2 (right): A student-made model from Figure 1 converted into a network of interconnected concepts.

How can network approaches be used to assess learning evident in models made by students in biology and statistics?

This study will apply network analysis to describe changes in students' model construction across different educational contexts as learners progress in introductory biology or statistics courses. We have accrued a large database of students' model-based responses, including models collected from exams, quizzes, homework, and in-class activities at multiple time points during courses. Each model is associated with meta-data including course, semester, class size, instructor, assessment type, model type, prompt, and date of collection. While we have been able to convert models into network representations (Figure 2), more work is needed to understand the network metrics as indicators of the learning process. To convert student-made models into network representations, we treat concepts as nodes and the relationships as edges, indicators can be used to understand the patterns and relations underlying the student models. The indicators include the number of nodes and edges, the density of edges to nodes, the path length between nodes, clustering coefficients, and centrality measures. Additionally, these models can be aggregated across a class to provide a full network summary of class responses. In this way, instructors may be able to see where students are making connections, and where relationships between ideas could be strengthened, which will inform instruction at both an individual and class cohort level. In addition to applying network metrics, models can also be analyzed for changes. These models were collected from multiple timepoints from 1,600 students in biology and 67 students in statistics. By analyzing a time series of models within each course (figure 3), we can gain clarity about the kinds of knowledge restructuring that occurs among students. This is especially useful for seeing if specific concepts or connections are "rewiring" during classroom learning.

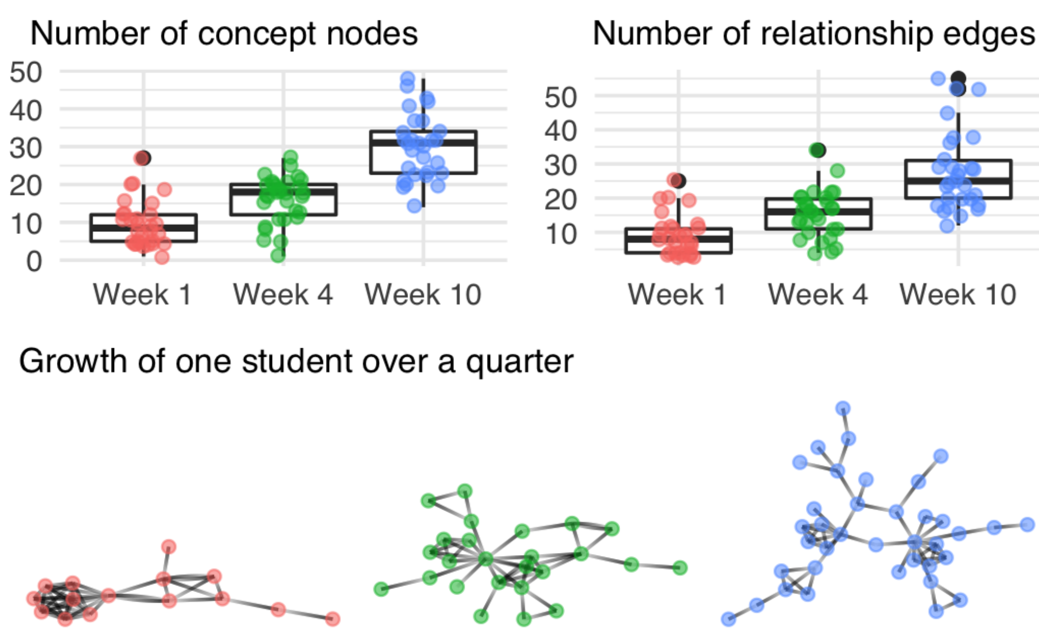


Figure 3. Preliminary data of changes to models made by statistics students throughout a quarter given the same prompt and wordbank, "Make a concept model to explain how statistics concepts are related." and a single student's progression across a course.

How can the simulations of network growth dynamics support an explanation of the learning processes evident in student models?

By relating these above-mentioned indicators, we can determine which ones are particularly indicative of major restructuring. Network theory suggests that networks can grow in different ways and each type of growth leaves distinct traces similar to the rings of a tree. Simulations of network growth dynamics produce signature characteristics. For instance, small-world networks grow from neighboring connections so these networks have short average path lengths and high clustering coefficients [10]. Analyzing student-made models will allow us to see if these growth patterns match the structures underlying the student models, which could provide insights into the cognitive restructuring processes occurring as students learn domain-specific knowledge in biology and statistics. These prospective findings will raise new questions to explore regarding the value of models for assessing and teaching, the types of patterns evident in student models, and the dynamic restructuring that occurs. As a consequence, this work aims to strengthen cross-institutional collaboration and lead to a sustained interdisciplinary research program.

Background

Theory of knowledge structures

Mental models reflect how one constructs and organizes knowledge about information and experiences [11], as well as how new knowledge is related to one's existing cognitive structure [12]. As an internal construct, students' mental models are not readily accessible via any direct assessment, but external representations in the form of behaviors, verbalizations, visualizations, and models can reveal select aspects of mental models and corresponding knowledge structures. When an individual generates a representation intended to match a mental model, the contents and structure of the representation can be used to infer information about knowledge structure – the specific, organized collective of concepts and relationships associated with a particular topic or domain [13,14]. As situations or contexts change, individuals must adjust their mental models, including which knowledge structures are accessed and how they are arranged. As such, model-based representations may be particularly useful for communicating the reorganization of knowledge structures in ways that are difficult to express in written and spoken language, which is serial.

Network theory

Like other networks found in biology, on the Internet, and in energy grids, concept models tend to exhibit patterns in their network structure [8,9,15]. Network graphs are composed of a set of nodes and a set of edges that connect nodes. Researchers have applied network metrics (e.g., surface structure, connectedness, ruggedness, etc) derived from graph theory to describe students' cognitive structures represented in concept maps and models [12]. As a previous result from part of this dataset, Dauer *et al.* used a network attribute related to connectedness to describe changes in biology models over time, which suggested that students were first accreting knowledge and then pruning relationships to produce more parsimonious models [16].

C. Procedures

Project team

Primary PI: Caleb Trujillo, Assistant Professor, School of Interdisciplinary Arts and Sciences, University of Washington Bothell. calebtru@uw.edu. Dr. Trujillo will manage data collection of models made at UW, perform analysis for network analysis, seek external fundings, mentor undergraduate students, and report findings in a publication when appropriate.

Collaborators: Tammy Long, Associate Professor, Department of Plant Biology, Michigan State University. longta@msu.edu. Dr. Long will continue to manage the database of models made at MSU, contribute to reporting of findings, and provide supervision and collaboration for applying for external funding. As Trujillo's former Principal Investigator, Dr. Long can provide continuity in faculty development.

Undergraduate research assistants at UW-Bothell (to be identified) will assist with data collection, data entry, analysis, and reporting for the full year of funding.

Phases

Phase 1: Preprocessing and transcription. Received IRB approval from UW. About 3,500 from MSU and all models from UWB have been transcribed into the digital format. Transcription of student models from a pen-and-paper form to a digital form will require ongoing work from undergraduate researchers as more models become available.

Phase 2. Workflow and database development. We have streamlined a previous workflow to import the digital model into a database to store as a network representation. The conversion of these models into a network format are written in R[17] and tracked using GitLab or GitHub.

Phase 3. Analysis and visualization. We are currently working on this phase. After creating the network format, we are now deciding which indicators to calculate for each model and aggregating across models. Additionally, these data will be visualized in Shiny and Gephi [18,19].

Phase 4. Simulating network growth dynamic. Once results have been collected from the student-made models, we will compare the indicators to networks generated by a computer using different growth dynamic algorithms.

Phase 5. Dissemination and growing research. We will report our results at conferences (e.g. Society for the Advancement of Biology Education, Gordon Research Conference - Visualization in Science and Education) and in manuscripts submitted for publication to science education research journals such as *CBE-Life Science Education*. Additionally, we will apply for external funding using the preliminary results from the National Science Foundation for programs such as Transdisciplinary Research in Principles of Data Science (TRIPODS) and Building Community and Capacity in Data-Intensive Research in Education (BCC-EHR).

With great appreciation from the UWB Scholarly Research and Creative Practice (SCRP) seed grant, we completed the first and second phases between Summer 2020 - Summer 2021 and are currently working on phase 3 of 5.

D. Time schedule

Phase	`20	`21	'22				`23	
	SRCP Sum. -Sum.	Aut	Win.	Spg.	Sum.	Aut	Win.	Spg.
1. IRB, Preprocessing, Transcription of models	Complete		X	X			X	X
2. Workflow, database development	Complete							
3. Analysis, visualization	In progress	X	X	X	X	X		
4. Simulating growth dynamics			X	X	X	X	X	
5. Dissemination. Submit work to CBE-LSE, SABER, GRC-Vis. Apply for NSF grants.					X	X	X	X

E. Need for RRF Support

Alignment with RRF mission

This project aligns with the mission of RRF by supporting faculty who are junior in rank and providing unique opportunities to increase applicants' competitiveness for subsequent funding.

If awarded will help grow a new approach to assessment in science education that has been initially seeded by an internal grant at UWB. With great appreciation from the UWB Scholarly Research and Creative Practice (SCRP) seed grant, we completed the first and second phases between Summer 2020 - Summer 2021 and are currently working on phase 3 of 5. This foundational work has provided some preliminary data but more analysis is needed to show the depth of this approach in order to be competitive for NSF funding. An early version of this application was sent to NSF EHR Core: Leveraging Big Data to Advance Biology Education Research: A Cognitive Analysis of Learners' System Models (NSF-15-509), but was not funded.

This award will provide funds for junior faculty (Trujillo) that will complement existing start up funds supporting this research.

Project impact

The project will further an understanding of models used in STEM education and strengthen interdisciplinary connections between disparate fields such as data science, learning theory, and discipline-specific education. In doing this, our research can provide evidence that fulfills recommendations from leaders on policy regarding educational changes [1,2,5,7].

Supplemental materials

- Budget
- Curriculum Vita
- Research support
- References

4. Budget pages

Itemized budget and budget narrative

Personnel (01)	**$28,134**	
Caleb Trujillo, PI (2 mos effort SUMR)		$17,115
TBD, Research Assistant (390 hours SUMR)		$8,550
TBD, Research Assistant (145 hours ACAD)		$2,469
Fringe Benefits (07)	**$6,322**	
Caleb Trujillo (23.2%)		$3,971
TBD, Research Assistant (21.6%)		$1,847
TBD, Research Assistant (20.4%)		$504
Total Project Costs	**$34,456**	

Trujillo will mentor an undergraduate and/or research assistant on phases 1-4. Undergraduate researchers will work to transcribe the models into a digital format and learn how to conduct a network analysis in R.

Additionally, the student will transcribe new models made from the classes across the quarters that teach statistics.

5. Curriculum vitae

NSF Biographical Sketch

Caleb M. Trujillo

18115 Campus Way NE, UW2-224
School of Interdisciplinary Arts and Science, University of Washington Bothell
Bothell, WA 98011
email: calebtru@uw.edu
425-352-3317

(a) Professional Preparation

University of Colorado, Boulder, CO - Molecular, Cell., & Dev. Biology	B.A.	2009
Purdue University, W. Lafayette, IN - Biological Sciences	Ph.D.	2012
Michigan State University, E. Lansing, MI - Biology Education	Postdoc	2015–2019

(b) Appointments

University of Washington Bothell. Bothell, WA
 Assistant Professor, School of Interdisciplinary Arts and Science 2019–Present

Michigan State University. E. Lansing, MI
 Research Associate, College of Natural Science 2017–2019
 Research Associate, Department of Plant Biology 2015–2018

Purdue University. W. Lafayette, IN
 Research Assistant, College of Science 2011–2013

University of Colorado at Boulder. Boulder, CO
 Professional Research Assistant, Carl Wieman Sci. Educ. Initiative 2009–2010
 Temporary Research Assistant, MCDB Department 2009–2010

(c) Products

(i) Relevant Publications

Nguyen, H.N., Trujillo, C., Wee, K. Bowe, K. (2021) Interactive qualitative data visualization for educational assessment. *Association for Computing Machinery (ACM) International Conference Proceedings*.

Trujillo, C.M., & Long, T. M. (2018) Document co-citation analysis to enhance transdisciplinary research. *Science Advances, 4*(1), e1701130

Trujillo, C.M., Anderson, T.R., & Pelaez, N.J. (2016). An instructional design process based on expert knowledge for teaching students how mechanisms are explained. Advances in *Physiology Education, 40*(2), 265-273.

Trujillo, C.M., Anderson, T.R., & Pelaez, N.J. (2015). A model of how different biology experts explain molecular and cellular mechanisms. *CBE- Life Science Education, 14*(2), 1-13.

Perez, K.E., Hiatt, A., Davis, G.K., Trujillo, C., French, D.P., Terry, M., & Price, R.M. (2013). The EvoDevoCI: A concept inventory for gauging students' understanding of evolutionary developmental biology. *CBE- Life Science Education, 12*(4), 665-675.

(d) Synergistic Activities

Innovations in teaching and training: Designing assessments of teacher knowledge Engaging Educators in Developing and Using Molecular Case Studies at the Interface of Biology and Chemistry. National Science Foundation (NSF) RCN UBE award to Dutta (PI), Trujillo (Co-PI).

Innovations in teaching and training: Published a classroom lesson for teaching large enrollment biology courses about data collection and visualization through measuring seasonal change of campus trees. Emery, N.C., Trujillo, C.M., Jarosz, A., & Long, T.M. (2019*).* Quantifying and Visualizing Campus Tree Phenology. *Course Source.*

Contributions to the science of learning: Recipient of the 2018 Postdoctoral Excellence in Research Award at Michigan State University from the Graduate School for contributions to science education research.

Service to the scientific and engineering community outside of the individual's immediate organization: A peer reviewer for multiple journals including *Science Advances; BioScience; CBE – Life Science Education; Advances in Physiology Education; Sustainability; Journal of College Science Teaching.*

Broadening the participation of groups underrepresented in STEM: Elected Vice President and Board Member of Purdue Chapter Society for Advancement of Chicanos and Native Americans in Science (SACNAS) which was awarded Outstanding Leadership on Campus 2012 and Outstanding Governance Award 2014.

6. Research support

Funded projects

Engaging Educators in Developing and Using Molecular Case Studies at the Interface of Biology and Chemistry. National Science Foundation (NSF) award to Dutta (PI), Trujillo (Co-PI).	2020-2025 $500,000 NSF RCN-UBE 2018884
Teaching and Leadership Innovation Initiative: Instituting Integrated Statistics Learning Environment (ISLE). Award to Trujillo (PI) and Ferrare (PI).	2020-2021 $3000 from UWB
Student-made Models Analyzed on a Large Scale (SMALS). Funding provided through UW Bothell Scholarship, Research, and Creative Practice (SRCP). Award to Trujillo (PI).	2020-2021 $18,597 from UW
Visualizing qualitative data for science and education. Funding provided through Gordon Research Conference Visionary Award to Jeffery, Nguyen, Trujillo, Wee (PIs).	2019-2021 $9,994 from NASA

Start up

Total start up funds include an initial $25,000 and a $5,000 extension. The remaining budget is $26,742 and is allocated for the following expenses:

- $4,000 for travel to conferences and collaborators (when safe to do so)
- $1,000 for server development and maintenance
- $1,000 for software, subscriptions, and transcription services.
- $15,000 for personnel (I currently have two post-bacc research assistants and two student researchers, so this full amount is budgeted for them across multiple projects).
- $4,900 for computer equipment, and technology for personnel (pending)

While there is a significant amount of start-up remaining, this is due to (1) being unable to travel, (2) expenses being applied to the SRCP grant, (3) an extension awarded by the school of IAS, and (4) not yet being able to create a physical research lab in a shared space during the Covid-19 pandemic since Trujillo started employment Sept. 2019, 5 months before lock-down. The requested funds will allow personnel to work on this research project and provide the time to write research reports, mentor, and apply for NSF grants.

7. Literature References

1. Bhatt, C., Sajja, P. S. & Liyanage, S. *Utilizing Educational Data Mining Techniques for Improved Learning: Emerging Research and Opportunities: Emerging Research and Opportunities*. (IGI Global, 2019).
2. President's Council of Advisors on Science and Technology (U.S.). *Report to the President, Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*. (2012).
3. Moharreri, K., Ha, M. & Nehm, R. H. EvoGrader: an online formative assessment tool for automatically evaluating written evolutionary explanations. *Evolution: Education and Outreach* vol. 7 (2014).
4. Urban-Lurain, M. *et al.* Beyond multiple choice exams: Using computerized lexical analysis to understand students' conceptual reasoning in STEM disciplines. *2009 39th IEEE Frontiers in Education Conference* (2009) doi:10.1109/fie.2009.5350596.
5. National Research Council, Division of Behavioral and Social Sciences and Education, Board on Science Education & Committee on the Status, Contributions, and Future Directions of Discipline-Based Education Research. *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*. (National Academies Press, 2012).
6. Handelsman, J., Miller, S. & Pfund, C. *Scientific Teaching*. (Macmillan, 2007).
7. States, N. L. *Next Generation Science Standards: For States, By States*. (National Academies Press, 2013).
8. Börner, K., Sanyal, S. & Vespignani, A. Network science. *Annual Review of Information Science and Technology* vol. 41 537–607 (2007).
9. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
10. Steyvers, M. & Tenenbaum, J. B. The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science* vol. 29 41–78 (2005).
11. Johnson-Laird, P. N. Mental Models in Cognitive Science. *Cognitive Science* vol. 4 71–115 (1980).
12. Ifenthaler, D., Masduki, I. & Seel, N. M. The mystery of cognitive structure and how we can detect it: tracking the development of cognitive structures over time. *Instructional Science* vol. 39 41–61 (2011).
13. Clariana, R. B. Deriving Individual and Group Knowledge Structure from Network Diagrams and from Essays. *Computer-Based Diagnostics and Systematic Analysis of Knowledge* 117–130 (2010) doi:10.1007/978-1-4419-5662-0_7.
14. Gobbo, C. & Chi, M. How knowledge is structured and used by expert and novice children. *Cognitive Development* vol. 1 221–237 (1986).
15. Barabási, A.-L. & Albert, R. Emergence of Scaling in Random Networks. *Science* vol. 286 509–512 (1999).
16. Dauer, J. T., Momsen, J. L., Speth, E. B., Makohon-Moore, S. C. & Long, T. M. Analyzing change in students' gene-to-evolution models in college-level introductory biology. *Journal of Research in Science Teaching* vol. 50 639–659 (2013).
17. R Development Core Team. *The R Reference Manual: Base Package*. (Network Theory., 2003).
18. Hadley, M. J. *Building Data Apps with R and Shiny: Essential Training*. (2019).
19. Cherven, K. *Network Graph Analysis and Visualization with Gephi*. (Packt Publishing Ltd, 2013).