Three Applications of Semantic Network Analysis to Individual Student Think-Aloud Data

Jennifer G. Cromley[a]*, Joseph F. Mirabelli[b], and Andrea J. Kunze[c]

a Department of Educational Psychology, College of Education, University of Illinois at Urbana-Champaign

b Department of Biomedical Engineering, University of Michigan Ann Arbor

c Department of Psychology, Butler University

Revision submitted for review on October 6, 2024 to *Contemporary Educational Psychology*

*Corresponding author

This material is based upon work supported by the National Science Foundation under Award No. 2225298. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Abstract

Student during-learning data such as think-alouds or writing are often coded for use of strategies or moves, but less often for what knowledge the student is using. However, analyzing the *content* of such products could yield much valuable information. A promising technique for analyzing the content of student products is semantic network analysis, more widely used in political science, communication, information science, and some other social science disciplines. We reviewed the small literature on semantic network analysis (SemNA) of individuals with relevant outcomes to identify which network analysis metrics might be suitable. The Knowledge Integration (KI) framework from science education is discussed as focusing on amount and structure of student knowledge, and therefore especially relevant for testing with SemNA metrics. We then re-analyze three published think-aloud data sets from undergraduate students learning introductory biology with the metrics found in the literature review. Significant relations with posttest comprehension score are found for number of nodes and edges; degree and betweenness centrality; diameter, and mean distance. Inconsistent results possibly due to text-specific features were found for number of clusters, LCC, and density, and null results were found for PageRank centrality and centralization degree. Basic principles from the KI framework are supported—amount of information (nodes), connections (edges, average degree), key ideas (degree and betweenness centrality) and length of causal chains (mean distance and diameter) are related to posttest comprehension, but not density or LCC. Possible explanations for slight variations across data sets are discussed, and alternative theories and metrics are offered.

Keywords: causal chains; during-learning data; educationally-relevant outcomes; propositional analysis

Three Applications of Semantic Network Analysis to Individual Student Think-Aloud Data

1. Introduction

Education researchers quite often collect verbal data to try to measure student knowledge during learning, such as think-alouds or discourse, student writing, or brief responses in online learning environments, which we will refer to as student products. Student products are often coded for use of strategies or moves, but less often analyzed to measure what knowledge the student is using. That is, student products are typically coded for strategic operations such as summarizing or making inferences, activating background knowledge, and so on. A few techniques are known for quantitatively measuring knowledge structures (e.g., card sorts used with multidimensional scaling [Irby et al., 2016], laboratory categorical priming studies), but not for during-learning data collected in real time.[1] A promising technique for measuring characteristics of the *content* of student products is semantic network analysis (SemNA), more widely used in political science, communication, information science, and some other social science disciplines (Scott, 2017).

What characteristics of the content of student products might be informative in educational psychology? At the simplest level, we can count how many facts students know

[1] Note that we are excluding 4 types of semantic network analysis that could be familiar to some readers:
1) *associative networks* such as single vocabulary words, where reaction time is used to map how closely the meaning of one word is associated with the meaning of another word (e.g., Siew, 2022).
2) *epistemic network analysis* which uses automated analyses of co-occurrence of words to create a visual network representation and one metric called the centroid (e.g., Wooldridge et al., 2018).
3) *pathfinder networks* which are a psychometric technique usually applied to similarity ratings (e.g., Clariana & Koul, 2008)
4) *questionnaire-based semantic networks* which use Likert-type responses to questionnaire items to infer a structure of beliefs (e.g., Fishman & Davis, 2022)

(called nodes in SemNA) and how many other facts each of those facts are connected to (called edges in SemNA), which as a set comprise that person's knowledge network. These concepts are borrowed from social network analysis, where we can quantify each person's (node) relationships (edges) in some community of practice such as a classroom, learning environment, workplace, neighborhood, and so on. At a more complex level, we can quantify the inter-connectedness of information in a whole knowledge network in at least a couple of ways—how many connected nodes each node is connected to (distance) and how many connections are present out of the total possible (density).

Why might measuring these aspects of a knowledge network be useful? We begin by summarizing a popular science education framework that describes changes in the characteristics of knowledge networks with development—the Knowledge Integration (KI) framework (Linn, 2006). We then review which network metrics (network analysis measures) have been used for SemNA related to learning-relevant outcomes such as grades or expertise. We then present three re-analyses of our own think-aloud data to show the value of SemNA metrics as measures that can then explain variance in post-reading comprehension.

1.1. The Knowledge Integration (KI) Framework

The Knowledge Integration (KI) framework of Linn (2006) and colleagues is rooted in decades of research showing that students often pick up isolated facts in science classes, from experiences in daily life, and in many informal learning contexts. With development and/or specific KI instruction, students both gain more factual knowledge, and that knowledge becomes more richly interconnected. One component of KI instruction is presenting canonical scientific ideas, to which students can—and do—link some previously-isolated facts. KI is a specific expertise theory (cf., Ericsson et al., 2018) in that the size and structure (specifically, the interconnectedness) of the knowledge base grows with development. In the KI framework, students can have varied approaches to holding contradictory ideas, but with development, students on average not only come to know more about scientific topics, but their knowledge becomes more densely linked. With instruction that encourages students to gather evidence and reason with it, students can retain accurate information and form highly-interconnected, consistent knowledge structures to replace the sparse and disconnected knowledge they began with. KI instruction can include different components, but there is an emphasis on gathering data, making predictions, comparing alternative explanations, reflecting on results, and working collaboratively. These elements of KI instruction result in students not only knowing more concepts, but also connecting those concepts to a greater degree and in scientifically accurate ways, and thereby being better able to solve novel problems.

Both conventional classroom science instruction and conventional science assessment items tend to focus on factual learning, with few interconnections among facts (Clark & Linn, 2003). KI instruction, by contrast, focuses on making such connections. Conventional science lessons tend to focus on structures such as the different parts of a flower, with much less focus on how the parts function together or as part of a system (e.g., insects transferring pollen from the anther of one plant to the stigma of another plant in the same species in order to fertilize an ovary in the second plant; Hmelo-Silver & Pfeffer, 2004). In biology learning, this especially undermines the development of a core concept in biology that function (e.g., pollen being easily moved from one plant to another) depends on structure (e.g., pollen being a loose dust; Kohn et al., 2018). When teachers are overly focused on teaching structure for the sake of knowing structure—that is, a focus on facts—students are especially likely to a) miss the structure-function connections and b) memorize elements but not relations.

Overall, the KI framework has 5 specific implications for the network properties of student knowledge in such fact-heavy classrooms, in that students are likely to have 1) many nodes but few edges connecting nodes and consequently 2) more isolates and 3) lower network density, 4) edges that are more likely to be definitions or ‘part-of’ rather than ‘acts-on’ edges, and as a consequence, 5) there may be shorter causal chains in these networks (see Figure 1 for two contrasting examples from one of our data sets).

Fig. 1. *Two contrasting knowledge structures—graphs from think-aloud-data*

With development or targeted KI instruction, by contrast, students with a better understanding of science are likely to focus on structure-function connections and learn elements and relations together. This has the opposite implications for the network properties of student knowledge according to the KI framework, in that expert students are likely to have 1) many nodes and many edges connecting nodes and consequently 2) fewer isolates (which implies fewer clusters), 3) higher network density, 4) edges that are more likely to be causal action verbs, and as a consequence, 5) there may be longer causal chains in these networks. Furthermore, 6) when KI instruction introduces important scientific ideas (such as a substance boiling when heat adds sufficient energy to separate molecules, rather than isolated facts such as water boils at 100° C), then those important ideas should be central in student networks. Indeed, KI researchers (Schwendimann, 2014; Schwendimann & Linn, 2016) found that after KI instruction there were more nodes and links, higher degree centrality for important scientific ideas, and higher network density in students’ concept maps than before KI instruction, as well use of a higher conceptual level for edge (verb) types assigned weights using a KI rubric, all using a KI-specific weighting scheme that weighted edges less when subject-verb-object phrases captured relations of low importance (e.g., the example of water→100° C above) vs correct *causal* relations.

In summary, the KI framework predicts that for students with a better understanding of science, their networks should have more nodes and more edges (which implies higher density, many triads, and less variability in edges/node), the nodes should be connected in long causal chains (implying higher mean distance [which represents cumulated node-level closeness centrality], and diameter, smaller number of clusters/fewer isolates, and therefore a large cluster that includes most of the nodes); and key scientific ideas should have high centrality, both highly connected (degree centrality) and including causes-of and effects-of (betweenness centrality).

1.2. Literature Review of SemNA Methods Applied to Individual Student During-Learning Data and Related to a Learning Outcome

Network analysis broadly defined looks at people/ideas, how they are connected to each other in a network, and how properties of the network can affect various human activities. Social network analysis focuses on how people are connected (e.g., by collaborating, conversing, helping, and so on), the properties of those social networks, and how network properties affect the people in the network. Another network analysis method, semantic network analysis (SemNA), focuses on verbal data, how ideas are connected to each other (e.g., by co-occurrences of terms, by verbs, by repetitions), and what the structure of the knowledge can reveal about the people who produced the verbal data. In this literature review, we focus on a subset of SemNA approaches that would be applicable to relating the measured characteristics of student products to what the student learned (see Figure 2). Therefore, we sought SemNA studies reporting on analyses of metrics applied to individual people's products (not only the products of a whole class), where the focus was on the content of the student product rather than the social spread of ideas, where content knowledge was measured rather than smaller units such as oral vocabulary words, and where network properties were related to a learning outcome such as a test score, grade, level of expertise, questionnaire score, and so on. Note that in all cases, scientifically accurate information (nouns) was included in the semantic networks to which SemNA metrics were applied.

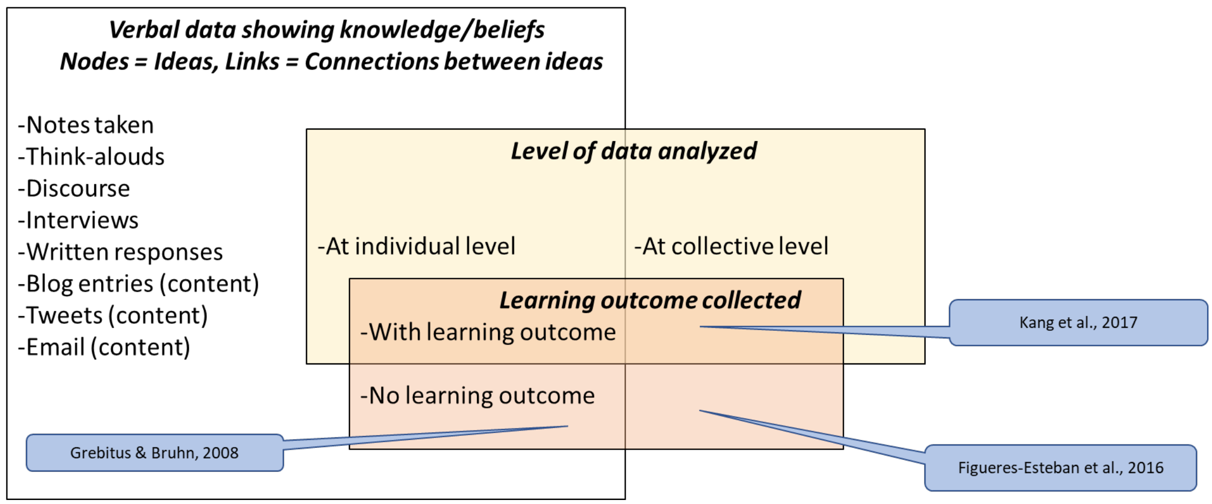


Fig. 2 *Schematic of verbal during-learning data at individual level related to learning outcomes. Note:* The references are examples of excluded studies

The purpose of this literature review is twofold: first, our purpose is to share a technique that others have applied so that we can document the work of scholars who have discovered how to do SemNA with individual student products. Our second purpose is to document which of dozens of existing network metrics have been shown in at least 2 studies to relate to learning when individuals' knowledge networks are analyzed. Note that different SemNA metrics can capture quite similar phenomena, so results from different SemNA metrics might be quite similar (Morrison et al., 2022). A metric used in only one SemNA study was judged to be too idiosyncratic to warrant applying it to our think-aloud data. Our review seeks to characterize a very disconnected literature, one where few of the studies we identified cited more than 2 other studies we identified, and where very different terminology was used to describe similar SemNA techniques.

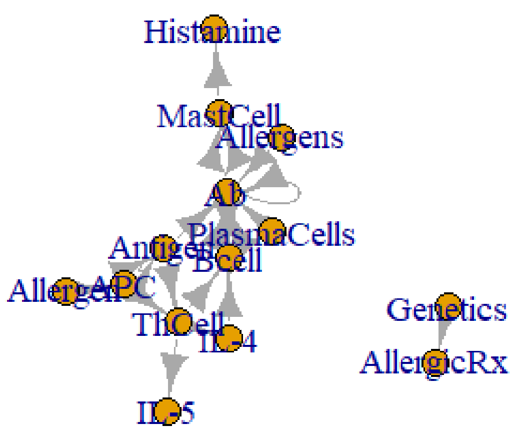
The starting point of our search was Siew's (2022) review, in which she advocated for more SemNA in education research. From the candidate articles she cited, we searched by keywords (e.g., "Network Analysis" as a subject term in the ERIC and Academic Search Ultimate databases), author names, and cited article searches for published papers and conference proceedings that fit the criteria above and identified 32 empirical studies applying SemNA to individual data with learning outcomes (see Supplementary Table 1 for the complete list of studies, details of samples, tasks, and research design, and which metrics were used in them). Learning outcomes included researcher-scored posttests, participants' categorization as novices vs. experts, changes in the semantic network before vs. after instruction, scores on standardized questionnaires, assignment or course grades, or performance on a real-life task.

These 32 studies spanned driving safety (Salmon et al., 2013), health and mental health learning (e.g., causes of obesity, Frerichs et al., 2018); mindfulness meditation (Pokorny et al 2018), learning to become an entrepreneur (Laukkanen, 2023), introductory psychology (Siew, 2019), as well as mathematics (e.g., concepts of triangles; Haiyue & Yoong, 2010), science (e.g., chemistry, Podschuweit & Bernholt, 2020; physics, Bodin, 2012), and other typical school subjects. In addition to introducing each metric in the following review, we reiterate how each metric would be expected to relate to scientific understanding for students with a better understanding of science in the KI framework.

Note that metrics are not statistics, they are more like a measure or a way of quantitatively measuring characteristics of a node or network; they are not tested against any underlying distribution so they are not 'significant' or 'non-significant' in and of themselves.

1.2.1. Nodes

Each idea in a network is called a node (these are represented as circles in an example plot shown in Figure 3). In the literature, number of nodes correlated positively with achievement (Freedman et al., 2024, Study 1; Hoppe et al., 2012; Kim, 2024; Wei & Yue, 2016). Experts included more nodes in their concept maps than did novices (Gogus et al., 2009; Kapuza et al., 2020; Smith & Parrott, 2013; Wagner, et al., 2020; Wagner & Priemer, 2023). Number of nodes significantly increased over time (Ifenthaler et al., 2011; Laukkanen, 2023) and after instruction (Bodin, 2012; Dauer et al., 2019; Frerichs et al., 2018; Giabbanelli & Tawfik, 2021; Kapuza et al., 2020; Kim & McCarthy, 2021a). However, null results were found for number of nodes by Gobbo and Chi (1986), Kim and McCarthy (2021b), Podschuweit and Bernholt (2020), Siew (2019), and Zhou et al. (2015).



# of Nodes = 14
# of Edges = 23
In-Degree for Antibody = 5
Out-Degree for Antibody = 3
In-Degree for Histamine = 1
Out-Degree for Histamine = 0
In-Degree for Phagocytes = 2
Out-Degree for Phagocytes = 1
Diameter. = 5
Mean distance. = 2.17
Components = 2
Density = .13
Centralization degree = .23
LCC Percent = .86

Fig. 3 Sample plot of a network with its accompanying metrics

In 15 of 21 studies we identified that used the number of nodes metric, that metric was positively related to a learning outcome. Note that non-significant results could be due to a true absence of a relation in the population, using an overly-small sample size, or other characteristics of the research (e.g., learner-task match, poor fidelity of intervention). Therefore any single instance of non-significance summarized here needs to be considered in context. In the KI framework, one would expect that more nodes may be found among students with a better understanding of science, provided those nodes are also densely connected.

1.2.2. Edges

The raw count of connections between nodes is referred to as the number of edges (these are represented as arrows in an example plot in Figure 3). In the literature, number of edges was positively correlated with achievement (Freedman et al., 2024, Study 1), Guerra et al., 2015; Hoppe et al., 2012; Wei & Yue, 2016; Yang et al., 2018; Zhou et al., 2015). Experts included more edges than did novices (Gogus et al., 2009), and more knowledgeable participants included more edges (Gobbo & Chi, 1986). Number of edges significantly increased over time (Laukkanen, 2023) and after instruction (Bodin, 2012; Dauer et al., 2019; Frerichs et al., 2018; Giabbanelli & Tawfik, 2021; Kim & McCarthy, 2021a). Null results were reported in Freedman et al. (2024, Study 2) and Kim and McCarthy (2021b). In 14 of 16 studies we identified that used the number of edges metric, that metric was positively related to a learning outcome. In the KI framework, where interconnectedness linking information is key, one would expect the number of edges to be positively associated with learning outcome.

1.2.3. Edges/Nodes Ratio, Average Degree

Raw count of edges is by definition always related to number of nodes, so an edge/nodes ratio (called average degree) can correct for this inflation when there are more nodes. Average degree is positively correlated with achievement (Freedman et al., 2024, Study 1; Hähnlein & Pirnay-Dummer, 2024; Hoppe et al., 2012; Ifenthaler et al., 2011; Kapuza et al., 2020; Kim, 2024; Wei & Yue, 2016; Yang et al., 2017, 2018). Experts have higher average degree than novices (Kapuza et al., 2020). Average degree is higher with intervention (Giabbanelli & Tawfik, 2021; Kim & McCarthy, 2021a; Lehmann et al., 2020). Null results were reported by

Freedman et al. (2024, Study 2), Guerra et al. (2015), Kim and McCarthy (2021b), and in some samples by Wei and Yue (2016). In 11 of 14 studies we identified that used average degree, that metric was positively related to a learning outcome. Similarly, in the KI framework, networks are expected to be both include more nodes and densely-connected nodes (i.e., more edges) for students with a better understanding of science.

1.2.4. Node-Level Metrics: Node Degree Centrality.

At the level of a single node, some nodes might be more connected than other nodes in a network, so certain conceptually 'key' nodes might have higher degree centrality (a count of all edges for each node). Certain key nodes had higher degree centrality for higher-scoring or higher-achieving students (Freedman et al., 2024, Study 1 and Study 2), Haiyue & Yoong, 2010; Hecking et al., 2017; Kubsch et al., 2019; Schwendimann, 2014; Wei & Yue, 2016; Yang et al., 2018). Degree centrality was higher for experts than novices (Gogus et al., 2009; Salmon et al., 2013) and was higher for high-knowledge participants (Gobbo & Chi, 1986; Smith & Parrott, 2013). Degree centrality increased over time (Laukkanen, 2023) and was highest for instructed vs. uninstructed concepts in intervention (Daems et al., 2014). However, degree centrality sometimes showed no relation with achievement (Wei & Yue, 2016) and showed both increases and decreases after intervention (Dauer et al., 2019). In 12 of 15 studies we identified that used average degree, that metric was positively related to a learning outcome. The KI framework anticipates students with a better understanding of science to have fewer isolated concepts, highly densely linked knowledge, and to include key concepts in the network, thus associating learning outcomes with higher degree centrality from during-learning networks.

1.2.5. Betweenness centrality

Certain nodes may sit between specific other pairs of nodes, just as a person in a social network could serve as a marriage broker or a conduit for news. Betweenness centrality captures the extent that a node links a pair of other nodes, whether the edges are one-way (directed) or are undirected connections such as co-occurrence of words. Betweenness is positively correlated with concept map scores (Freedman et al., 2024, Study 2, but not Study 1), and experts show higher betweenness for key concepts compared to novices (Wagner et al., 2020; Wagner & Priemer, 2023). Betweenness centrality was highest for terms that cut across 3 instructed concepts in an intervention (Daems, 2014). In 4 of 5 studies we identified that used betweenness centrality, that metric was positively related to a learning outcome. Causal relationships linking concepts are an important aspect of the KI framework, and causal chains in network graphs would increase betweenness centrality. The KI framework also suggests densely connected networks, which tend to have higher betweenness centrality.

1.2.6. PageRank Centrality

This metric has been key for the dominance of Google as a search engine and quantifies how much neighboring 'pages' are linked to other 'pages'. In a social network, two of one's own friends may be friends with each other, forming triads in the network. PageRank centrality will be high when there are many triads, but can be zero if there are no triads in the network. PageRank centrality increases for key nodes after instruction (Bodin, 2012). Pokorny et al. (2018), by contrast, found higher PageRank centrality for nodes related to a well-being questionnaire corresponded to lower questionnaire scores. In both of the studies we identified that used betweenness centrality, that metric was significantly related to a learning outcome but it was positive in once case and negative in another. Densely connected networks anticipated in the KI framework are more likely to include more triads of connected ideas; thus, PageRank centrality should be higher for students with a better understanding of science.

In summary, in 58 of 72 tests (81%), the node-level metrics (number of nodes, number of edges, edge/node ratio, and two centrality scores for key nodes [degree and betweenness]) show a pattern of positive relations to desirable educational outcomes. This is largely consistent with patterns in social network analysis though in semantic networks high-centrality nodes might not represent key people or ideas, but rather ideas that connect different topics that were instructed.

1.2.7. Network-Level Metrics: Diameter

Network diameter represents the longest path required to traverse the entire network. A shorter diameter is better in a social network and a longer diameter should be better in a knowledge network. Longer network diameter is associated with better achievement (Wei & Yue, 2016). Network diameter increases over time (Ifenthaler et al., 2011). Experts have sometimes been found to have a larger network diameter (Wagner et al., 2020), no difference (Wagner & Priemer, 2023), or a smaller network diameter (Salmon et al., 2013) than novices. Diameter increases with instruction (Giabbanelli & Tawfik, 2021). Null results were reported by Guerra et al. (2015) and by Kim and McCarthy (2021a, 2021b). In 6 of 9 studies we identified that used diameter, that metric was significantly related to a learning outcome, but in one case (expert and novice drivers) that relation was negative. Possibly this shows that novice drivers over-complicate the factors involved in the task—driving safely at grade-level railroad crossings—and hence produce longer causal chains when shorter once would be better. In the KI framework, graphs are expected to be highly dense for students with a better understanding of science, and densely connected networks should have higher diameters as there are fewer isolated, unconnected ideas.

1.2.8. Network density

Network density captures how interconnected all nodes are in a network, compared to a network with all possible interconnections. In some social networks every person could be connected to all other people, but in knowledge networks information might be locally interconnected or there can be long linear causal chains with (appropriately) few interconnections with nodes off that chain. Density is mostly positively related to achievement (Kubsch et al., 2019; Schwendimann, 2014; Yang et al., 2017), but a negative correlation was found by Freedman et al. (2024, Study 1). Experts have more dense networks than novices (Salmon et al., 2013). Density increases with instruction (Frerichs et al., 2018; Kim & McCarthy, 2021a, 2021b; Schwendimann, 2014). Null results were reported by Freedman et al. (2024, Study 2), Guerra et al. (2015) and Jamieson (2012). In 8 of 13 studies we identified that used density, that metric was positively related to a learning outcome. As previously mentioned, high density is expected for students with a good understanding of science based on the KI framework.

1.2.9. Mean distance

Mean distance captures the average number of nodes that can be traversed out from each node in the network. Longer mean distance is associated with higher quiz scores (Siew, 2019) and course grade (Guerra et al., 2015), but shorter mean distance is associated with higher concept map scores (Freedman et al., 2024, Study 1). Mean distance increases with instruction (Giabbanelli & Tawfik, 2021). Null results were found by Freedman et al. (2024, Study 2) and Kim and McCarthy (2021a, 2021b). In 3 of 6 studies we identified that used mean distance, that metric was positively related to a learning outcome. In a social network, a short mean distance is ideal, but in a knowledge network, a longer mean distance should result if there are long causal chains. In the KI framework, few isolates and well-connected networks are expected, and thus mean distance should be higher in general. However, one could expect that the domain being studied may determine the importance of mean distance (for example, biology learners read and hear about long causal chains when learning about biochemical processes; language learners may read about and hear smaller chains linking vocabulary words to definitions).

1.2.10. Network centralization

Network centralization is related to the standard deviation of the number of edges per node, across the whole network. High centrality means there is a single center of the whole network. Kim (2024) found that higher-comprehending participants had higher centralization on their concept maps. Kim and McCarthy (2021b) found that experts had lower centralization than novices, and also that in a writing course, centralization decreased over the semester. Wei et al. (2024) found higher delayed-posttest scores for a high-centralization concept map condition compared to a lower-centralization summary writing condition. However, Clariana et al. (2013) found no significant difference in network centralization between two different concept map conditions. In 3 of 4 studies we identified that used centralization, that metric was significantly related to a learning outcome, but with contradictory results (i.e., whether lower or higher centralization was found to be better). As previously stated, based on the KI framework we expect students with a better understanding of science to capture key ideas, thus we expect higher centralization would lead to better outcomes.

1.2.11. Number of clusters/components

The number of clusters (a.k.a. components), where each cluster is an interconnected network completely separate from all other clusters, should be an index of fragmentation or isolation, either of subgroups of people or of ideas. Yang et al. (2017; 2018) found that higher-achieving students had fewer clusters. Wagner and Priemer (2023) found that experts had fewer clusters in their concept maps than novices. However, Ifenthaler et al. (2011) found an increase in the number of clusters over a semester. In 3 of 4 studies we identified that used number of components, that metric was negatively related to a learning outcome (i.e., fewer clusters is better). One possible explanation for the number of clusters increasing in Ifenthaler et al. is that students first add disconnected knowledge when they learn new facts, and it tends to become interconnected only later. Based on the KI framework we expect a negative relation for this metric with science understanding, as high cluster count indicates more isolated information.

1.2.12. Largest Connected Component (LCC)

In many networks that have multiple components, there is one component that is much larger than the others. The LCC metrics concern the number of nodes in such a component; the LCC proportion captures the number of nodes in the LCC as a percent of all nodes in the network. Yang et al. (2017), and Yang et al. (2018) both found that larger LCC% was associated with achievement. However, Siew (2019) found no relation between LCC proportion in students' networks and their subsequent quiz scores. In 2 of 3 studies we identified that used LCC proportion, that metric was positively related to a learning outcome. Similarly, based on the KI framework, it is expected that most nodes for students with a better understanding of science would be part of a single, well-connected component, yielding a high LCC%.

Overall, the majority of tests of network metrics in the studies reviewed (25 of 39 tests; 64%) were significantly related to learning outcomes. Greater longest or mean distance, graph density, and LCC proportion are related to better learning, but graph centralization and number of components show mixed effects. In sum, the majority of studies reviewed suggest that larger and more strongly interconnected semantic knowledge networks having longer causal chains relate positively to better learning outcomes.

One consideration in choosing metrics is the co-occurrence of these metrics across this small literature. Across 36 studies we identified, number of nodes was used together with number of edges 13 times, with average degree 8 times, and with degree centrality and diameter 7 times each. Number of edges was used together with average degree 8 times and degree centrality 6 times. Diameter was used together with average degree 6 times. Other than these, there were no pairs of metrics used in common across studies in more than 5 out of 36 studies.

Node-level count metrics (number of nodes or edges and their ratio) tended to be used together with each other, and also together with degree centrality or diameter.

In sum, this literature review suggests that analyzing the content, and not just the 'moves,' revealed in verbal during-learning data such as think-alouds should have some explanatory power for posttest scores. We therefore applied SemNA metrics to three think-aloud data sets and tested the relation of these metrics to post-reading comprehension.

2. The current studies—overview, data processing, and data analysis

The main research question for the current studies was: How strong are the associations between during-reading knowledge structure characteristics measured in think-aloud data and post-reading comprehension? To answer this question, we applied network analysis (NA) metrics from literature to think-aloud data by re-analyzing three extant data sets. Study 1 reports on analyses of NA metrics applied to think-aloud data on a long textbook excerpt published in Cromley et al. (2020); study 2 reports on analyses of NA metrics applied to think-aloud data on two complementary 4-text sets using a within-subjects design published in Cromley et al. (2021). In all studies, participants provided informed consent and agreed to be audio recorded while thinking aloud (see below for details). Think-aloud directions neither listed nor modeled any reading strategies, and prompts comprised only "Say what you're thinking" or "Say what you're doing."

For all three data sets, audio recordings were transcribed verbatim, and after reading through the transcripts a master list of directed propositions (labeled Subject = Source, Verb, and Object = Target) was created using a common set of terms (see Supplementary Table 2 and a publicly-available how-to guide at https://hdl.handle.net/2142/121704). We then tallied propositions that students picked up on, defined as propositions that were included in any verbalization of passage content that was not reading or re-reading.

We then applied the 12 NA metrics identified from the literature review above to each participant's propositional network (see Supplementary Table 3 for R code using the *igraph* package). We then used Pearson correlations on rankit transformed data (Bishara & Hittner, 2012) to analyze the relation between each of these metrics and participant posttest comprehension score.

3. Study 1

3.1. Method

Study 1 is a re-analysis of think-aloud data from Cromley et al. (2020); please see that publication for more detail about the study.

3.1.1 Participants

Participants were 77 of the original 86 undergraduate biology course participants reported in Cromley et al. (2020); we selected participants who read three sequential pages from the original reading packet on the immune system. These participants did not differ from the sample in Cromley et al. by race, sex, or first-generation college status, but they had more recently completed the course we recruited from and had lower ACT reading and mathematics scores. They had a mean age of 20.1 (*SD* = 1.2), were 30% sophomores, 64% juniors, 5% seniors, and 1% post-baccalaureate. On average they had taken the introductory biology course 2.7 semesters (*SD* = 1.3) before participating. Seventy percent self-identified as female, 51% self-identified as White, 36% Asian, 4% Latine, and 9% of mixed or other race. Twenty-two percent had neither parent with a Bachelor's degree. Mean ACT reading scores were 30.1 (*SD* = 3.6) and mean ACT math scores were 30.9 (*SD* = 3.4).

3.1.2. Materials and procedure

In an individual session in our laboratory in Fall, 2016, participants gave informed consent including consent to be audiotaped and were asked to learn as much as they could from a set of provided textbook passages in pdf format in 40 minutes. The three consecutive textbook passages concerned mast cell immune response, the steps from antigen recognition to cytokine production inside cells, and antibody structure and function, and were taken in sequence from an introductory undergraduate biology textbook (Sadava et al., 2012).

Participants were given task instructions to think aloud while learning and were given paper and pen to take notes with the instruction: "if that is what you usually do when you are studying by yourself from your textbooks." After they were done reading, any notes were collected, and the audio recording was stopped, they were taken to another room and asked to type in a Word document everything they remembered from the text. Demographics were collected, and the process for payment of $35 compensation was explained.

3.1.4. Tallying propositions

Before tallying propositions from each set of think-aloud protocols, we read through all transcripts for participant content-related utterances that were not re-reading. The original transcripts had been formatted using *italics* for segments read (or re-read) from the text, and our focus was on other participant verbalizations related to the text set, either from text or diagram. Content-related verbalizations that we tallied were mostly in subject-verb-object form (noun/noun phrase followed by verb/verb phrase, followed by another noun/noun phrase), but lists of nouns—called isolates in SemNA—were also tallied to be consistent with KI theory. Only factually accurate verbalized propositions related to the text sets were tallied, but all legitimate synonyms were considered factually accurate (e.g., 'phagocytes gobble up the germs' for 'phagocytes engulf pathogens'). If a participant first verbalized an isolated fact but later connected the fact in a proposition within the same page (of the 4 pages in each text set), we tallied the proposition rather than the isolate. Non-content verbalizations that were not tallied included, for example, metacognitive monitoring, evaluating quality of text or figure, or noticing that there is more text or that there is a figure. Figure 4 shows a sample portion of one transcript with the tallied propositions. As can be seen in this figure, tallying was overall a low-inference process.

The allergen binds to IgE on a mast cell. Mast cells quickly release histamine, resulting in an allergic reaction. Ok. ***[Figure 2.** An Allergic Reaction.] An allergen is an antigen that stimulates B cells to make large amounts of IgE antibodies, which bind to mast cells and basophils. When the body encounters the allergen grain, these cells produce large amounts of histamine, which have harmful physiological effects.*	
[5:08] Alright. So this is about how allergens bind to B cells, then B cells clone of, *this causes a clone of the plasma cells to form*. So B cells get formed because there are multiple	Antigen binds Bcell
IgE antibodies and those antibodies are then connecting to the mast cell onto its receptors. And once the receptors connect to the	Ab binds MastCell
antibodies, then the mast cell releases a lot of histamine	MastCell releases histamine

Fig. 4 *Sample portion of one transcript (left) with the tallied propositions (right)*

Note: reading/rereading is shown in *Italics* and participant verbalizations are shown in plain text.

We used an iterative process to make the master list of propositions to tally from participant verbalizations by first making a list without tallying for any participant. We then began tallying propositions for participants, adding any new propositions to the master list. We continued this process, adding any new propositions, until we had created the master lists shown in Supplementary Table 2. The mean number of propositions tallied per participant was 28.7 (*SD* = 9.8).

3.1.5. Posttest scoring

In addition to detailed scoring of posttests explained in Cromley et al. (2020), the typed post-reading recalls were scored using researcher-developed concept maps that gave central concepts more credit (2 points) than supporting facts (1 point). All posttests were scored independently by two researchers.

3.1.6. Data analysis—choice of nodes for node-specific metrics

Inspection of the graphs and comparisons among a few each of high-scoring, medium-scoring, and low-scoring participants suggested we should systematically calculate node-level metrics for 3 terms: bacterium, cytokine, and phagocytes (note that these are the exact spellings we used). Not only did the node-level metrics suggest these 3 terms might be associated with learning, but they are conceptually important in the biology content covered in the text (bacteria trigger immune system cells to make cytokines which signal phagocytes [also immune system cells] which destroy bacteria) *and* they serve as the concepts which link the three passages.

3.2. Results

The main results of interest for Study 1 (as well as results for both Study 2 data sets) are shown in Table 2. The complete table of correlations, median and interquartile range across all metrics is found in Supplementary Table 4. Results of exploratory regressions are found in Supplementary Table 5.

Table 2

Pearson correlations with posttest comprehension by NA metric across all 3 data sets

NA Metric / *Data set*	*Study 1*	*Study 2 AH*	*Study 2 TC*
Number of nodes	.51*	.39*	.43*
Number of edges	.45*	.50*	.47*
Edge/node ratio	.41*	.43*	.23
Indegree	Bacterium: .25*	Antibody: .72*	Antigen: .44*
	Cytokine: .38*	Histamine: .26	MHCs: .n.c.
	Phagocytes: .33*	Mast cells: .39*	TCR: .31
Outdegree	Bacterium: .20	Antibody: .41*	Antigen: .47*
	Cytokine: .29*	Histamine: .23	MHCs: .47*
	Phagocytes: .26*	Mast cells: .33*	TCR: .32

NA Metric / *Data set*	*Study 1*	*Study 2 AH*	*Study 2 TC*
Betweenness	Bacterium: .28* Cytokine: .33* Phagocytes: .27*	Antibody: .49* Histamine: .34* Mast cells: .49*	Antigen: .49* MHCs: . n.c. TCR: .52*
PageRank	Bacterium: .03 Cytokine: .07 Phagocytes: .17	Antibody: .56* Histamine: .08 Mast cells: -.22	Antigen: .29 MHCs: -.23 TCR: .03
Density	-.34*	.10	-.16
Number of clusters	-.02	-.34	.04
LCC as %	.15	.37*	.27
Centralization degree	-.10	.05	.05
Mean distance	.41*	.42*	.53*
Diameter	.39*	.39*	.57*

Note: All reported correlations used the RANKIT percentile transformation (Bishara & Hittner, 2012), * indicates statistically significant at $p < .05$, LCC as % indicates number nodes in Largest Connected Component as percent of all nodes, n.c. indicates not calculable because specific edges were not present in the data set.

3.3. Brief discussion

Results for the NA metrics with comprehension of these 3-passage think-alouds were very consistent with significant findings in prior SemNA research on verbal data and were consistent with the KI framework for number of nodes, number of edges, and the edge/node ratio. Picking up more facts during learning and making more connections among them—even when controlling for the number of facts picked up via the average degree metric—yields significantly better post-reading comprehension scores.

Centrality results for comprehension on these think-alouds for the 3 key nodes—bacterium, cytokine, and phagocytes—were also quite consistent with prior SemNA research and were consistent with the KI framework for in-degree (causes of), outdegree (effects of; except for bacterium) and betweenness. However, results for PageRank centrality were non-significant; one possible explanation is the lack of triples (A→B, A→C, B→C [or C→B]) seen in the plots from this data set. Also, results for PageRank centrality were inconsistent in the two published studies we found (positive in Bodin, 2012 and negative in Pokorny et al., 2018). Picking up on more causes of and effects of key ideas in the text during learning is associated with better posttest comprehension. During-learning linking from and to (i.e., betweenness) key terms that connect the sequential passages is associated with better posttest comprehension.

Network-level results for comprehension on these think-alouds were very consistent for diameter and mean distance (correlated $r \sim .9$ with each other) with significant findings in prior SemNA research and were consistent with the KI framework. Verbalizing longer causal chains during learning—weather measured as the single longest chain in the entire network or the average chain length in the network—is associated with better posttest comprehension.

For this data set, network-level results for comprehension on these think-alouds were somewhat inconsistent for density, number of clusters, LCC as a percent of nodes, and centralization degree with findings in prior SemNA research and were somewhat inconsistent with the KI framework. We found density negatively related to posttest comprehension, whereas prior research had found positive relations, and KI research had found increases from pre- to posttest. We believe this is due to density being highly negatively correlated with number of nodes. This is negatively correlated because very small networks do tend to be more interconnected relative to the possible number of interconnections. We found no significant relation to posttest score for number of clusters, LCC as percent of nodes or for centralization degree. When there are many clusters, when the largest component captures a higher percent of all nodes, and when there is a larger SD for the number of edges per node, these are unrelated to posttest scores.

4. Study 2

4.1. Method

Study 2 is a re-analysis of think-aloud data from Cromley et al. (2021); please see that publication for more detail about the study.

4.1.1 Participants

Thirty undergraduate participants were recruited in fall 2018 after completing one of two introductory biology courses at a single US university. The majority were in their sophomore year (77%) and identified as female (80%). They were recruited via email and were compensated with a $35 gift card for participation.

4.1.2. Materials and procedure

The illustrated reading materials consisted of two 4-page text sets on the immune system, one of which described allergic hypersensitivity (AH) and the other the structure and function of t cell receptors (TC). The 4 pages within each set each came from a separate, reputable web or textbook source, and included overlapping, complementary information. None of the texts were the same as in Study 1, though there was substantial overlapping content between Study 1 and Study 2 texts. Each page had text on the left and one figure on the top right of the pdf page presented on a computer. Order of administration of the two text sets was counterbalanced across participants, but no differences were found for order of reading.

After consenting, participants were given think-aloud directions and were asked to read for the purpose of explaining the topic of the 4-page text set "as if you were explaining it to a peer." They were told they could read and think aloud about the 4 pages in any order and switch among pages as much as they wished. They were told that after the text was removed they would provide an oral explanation and a drawn explanation, which could be done simultaneously. No during-reading note-taking materials were provided. Participants read the first text set, provided their oral and drawn explanation of it, read the second text set, provided their oral and drawn explanation of it, and finally completed a demographics form. The entire session took about 1 hour.

4.1.4. Tallying propositions

Using the same procedures described for Study 1, we tallied from the think-aloud transcript each proposition the participant picked up on. The mean number of propositions per participant for AH was 10.0 (*SD* = 2.9) and for TC was 16.4 (*SD* = 6.3).

4.1.5. Posttest scoring

We typed up oral explanations for AH and TC to score these post-reading explanations. As explained in Cromley et al. (2021), we created a list of elements and relations for AH, then tallied the number of drawn AH elements, drawn AH relations, oral explanation AH elements and oral explanation AH relations. These 4 counts for AH were then z scored and combined into a single Principal Components score. We similarly created a list of elements and relations for TC, then tallied the number of drawn and oral explanation elements and relations. These 4 counts for TC were then z scored and combined in the same way using PCA. All posttest data were scored independently by two researchers.

4.1.6. Data analysis—choice of nodes for node-specific metrics

AH: Inspection of the graphs and comparisons among a subset of participants suggested we should calculate node-level metrics for 3 AH-specific terms: antibody, histamine, and Mast cell (note that these are the exact spellings we used). Not only did the node-level metrics suggest these 3 terms might be associated with learning, but they are conceptually important in the text set (mast cells are first responders to allergens, mast cells release histamine and also signal B cells to create antibodies specific to allergen proteins).

TC*:* Inspection of the graphs and comparisons among a subset of participants suggested we should calculate node-level metrics for 3 TC-specific terms: antigen, MHCs, and TCR (note that these are the exact spellings we used). Not only did the node-level metrics suggest these 3 terms might be associated with learning, but they are conceptually important in the text set (TCR [T cell receptors] recognize pathogenic or allergy-related antigen that is pushed to cell surfaces by the MHC molecule [Multiple Histocompatibility Complex]).

4.2. Results

The main results of interest for both Study 2 data sets are shown in Table 2. The complete tables of correlations, median and interquartile range across all metrics are found in Supplementary Table 4. Results of exploratory regressions are found in Supplementary Table 5.

4.3. Brief discussion

4.3.1. Text set AH

Results for the NA metrics with comprehension of these 4-page AH think-alouds were very consistent with significant findings in prior SemNA research on verbal data and were consistent with the KI framework for number of nodes, number of edges, and the edge/node ratio. As in Study 1, picking up more facts during learning and making more connections among them—even when controlling for the number of facts picked up via the average degree metric—yields better post-reading comprehension scores.

Centrality results for comprehension on these think-alouds for the 3 key AH nodes—antibody, histamine, and mast cells—were also quite consistent with prior SemNA research and were consistent with the KI framework for in-degree (causes of), outdegree (effects of; except for histamine) and betweenness. As in Study 1, results for PageRank centrality were mostly non-significant for comprehension (PageRank for Antibody showed quite a large relation at $r = .56^*$); we believe that the lack of triples—except for Antibody—explains this result. Picking up on more causes of and effects of key ideas in the text during learning is associated with better posttest comprehension. During-learning linking from and to (i.e., betweenness) key terms that connect the information—and are repeated across the complementary text sets—is associated with better posttest comprehension.

Network-level results for comprehension from these think-alouds in multiple complementary AH texts were very consistent for diameter and mean distance (correlated $r \sim .9$ with each other as in Study 1) with significant findings in prior SemNA research and were consistent with the KI framework. Making longer causal chains during learning—weather measured as the single longest chain in the entire network or the average chain length in the network—is associated with better posttest comprehension.

As in Study 1, network-level results for comprehension on these think-alouds from complementary AH texts were somewhat inconsistent for density, number of clusters, LCC as a percent of nodes, and centralization degree with findings in prior SemNA research and were somewhat inconsistent with the KI framework. We found density non-significantly related to posttest comprehension, whereas prior research had found positive relations, and KI research had found increases from pre- to posttest. Density in this data set is still negatively correlated with number of nodes but at $r = .6$. Again, this is negatively correlated because very small networks do tend to be more interconnected relative to the possible number of interconnections. We did not find a significant relation to posttest comprehension for number of clusters, but we did find a significant relation for LCC as percent of nodes. When reading the AH text set, having more nodes in the single largest connected cluster—but not having fewer clusters—is related to integrative comprehension. Centralization degree—a larger SD for the number of edges per node—is unrelated to integrative comprehension.

4.3.2. Text set TC

Results for the NA metrics with comprehension of these 4-page TC think-alouds were very consistent with significant findings in prior SemNA research on verbal data and quite consistent with the KI framework for number of nodes, number of edges, but not for the edge/node ratio. As in Study 1 and Study 2 AH, picking up more facts during learning and making more connections among them—even when controlling for the number of facts picked up via the average degree metric—yields better post-reading comprehension scores.

Centrality results for comprehension on these think-alouds for the 3 key TC nodes—antigen, MHCs, and TCR—were somewhat consistent with prior SemNA research and were somewhat consistent with the KI framework for in-degree (causes of; Antigen only), outdegree (effects of; Antigen and MHCs) and betweenness (Antigen and TCR only). As found in Study 1 and mostly in Study 2 AH, results for PageRank centrality were non-significant for comprehension; again, we believe that the lack of triples explains this result. Picking up on more causes of antigen and effects of antigen and MHCs—key ideas in the text—during learning is associated with better posttest comprehension. During-learning linking from and to (i.e., betweenness of) antigens and TCR, which are key terms that connect the information and are repeated across the complementary text sets is associated with better posttest comprehension. A possible explanation for finding fewer relations of centrality to comprehension in the TC set is the seeming structure focus of the text, which began with the first heading "Structure of the T cell receptor."

Network-level results for comprehension from these think-alouds in multiple complementary TC texts were very consistent for diameter and mean distance (weighted or unweighted; correlated $r \sim .9$ with each other as in Study 1 and Study 2 AH) with significant findings in prior SemNA research and were consistent with the KI framework. Making longer causal chains during learning—weather measured as the single longest chain in the entire network or the average chain length in the network—is associated with better posttest comprehension.

As in Study 1 and for the AH data set, network-level results for comprehension on these TC think-alouds were quite inconsistent for density, LCC as a percent of nodes, and centralization degree with findings in prior SemNA research and were somewhat inconsistent with the KI framework. We found density non-significantly related to posttest comprehension, whereas prior research had found positive relations, and KI research had found increases from pre- to posttest. Density in this data set is still negatively correlated with number of nodes but at $r = -.8$. Again, this is negatively correlated because very small networks do tend to be more interconnected relative to the possible number of interconnections. We did not find a significant relation to posttest score for number of clusters, LCC as percent of nodes, or for centralization degree. When reading the TC text set, having more clusters, having more nodes in the single largest connected cluster, or a larger SD for the number of edges per node is unrelated to posttest comprehension.

5. Extended discussion

How strong are the associations between during-reading knowledge structure characteristics measured in think-aloud data and post-reading comprehension? The answer depends on which knowledge structure characteristics as measured, as captured by different metrics. On the most-frequently used SemNA metrics from the literature and key elements of the KI framework—number of nodes, edges, and average degree—knowledge structure characteristics of our think-aloud data were significantly correlated with comprehension across 8 of 9 tests—picking up on more nodes and more edges during reading was significantly related to comprehension in our larger data set from 3 sequential texts with a typed posttest and in both analyses of the smaller data sets (AH and TC) from 4-page complementary multiple texts with a drawn-and-oral-explanation integrative posttest. Results for the edge-node ratio were significant across the first two data sets. Consistent with a strong positive trend in the literature, it appears that picking up more facts and connections during reading is associated with better post-reading comprehension, and researchers using think-alouds should consider this a useful metric. Based on two of three positive results for the edges/nodes metric, we also recommend its use for quantifying knowledge structures from think-aloud data.

On the degree centrality metrics from the literature and key elements of the KI framework—in-degree, out-degree, betweenness, and PageRank centrality—we found largely consistent results in analyses of our think-aloud data. It should be kept in mind that the chosen nodes in our three data sets—and indeed across the literature we reviewed—are specific to a text. For in-degree, we found 6 significant correlations out of 8 that could be calculated for in-degree; picking up on the cause(s) of key ideas in the text is associated with better post-reading comprehension, again across the different samples, sample sizes, texts, and posttest measures we used. For out-degree, we found 6 significant correlations out of 9 that were calculated for in-degree; picking up on the effect(s) of key ideas in the text is associated with better post-reading comprehension, again across the specific samples, sample sizes, texts, and posttest measures we used. For betweenness, we found 8 significant correlations out of 8 that could be calculated for betweenness; picking up on both cause(s) and effect(s) of key ideas in the text is associated with better post-reading comprehension, again across the specific samples, sample sizes, texts, and posttest measures we used. By contrast, PageRank centrality was only significant for 1 of 9 correlations we tested. One basis of these patterns of findings, we recommend that researchers who want to quantify knowledge structures from think-alouds should consider in-degree, out-degree, and betweenness to be useful metrics. PageRank might be a useful metric for analyzing semantic networks that have high inter-connectedness, such as might be found with end-of-year data tapping small numbers of nodes.

On the mean distance and diameter metrics from the literature and key elements of the KI framework we found completely consistent results in analyses of our think-aloud data. We also found the expected high correlation between mean distance and diameter ($r \sim .9$) in all data sets. We found 12 of 12 correlations for mean distance and diameter with comprehension were statistically significant, and these ranged from .39 to .67, which would be considered medium-large to large for education research (Cohen, 1988). Beyond linking pairs of nodes, making long chains of nodes is associated with better posttest comprehension. Interestingly, this is the opposite pattern from a 'good' social network, in which short mean distance and short diameter are associated with effective social networks (Watts & Strogatz, 1998).

On the other network-level metrics from the literature, one of which has been tested in the KI framework, we found almost no significant results in analyses of our think-aloud data. Density was negatively significantly correlated with comprehension in Study 1, and non-significant for both Study 2 data sets. Number of clusters was not significantly correlated with comprehension in any of our 3 data sets. Largest connected component nodes as a percent of all nodes was positively significantly correlated with comprehension in Study 2 for the AH text set only, and non-significant for the other two data sets. Centralization degree—a measure of the SD of node degree—was non-significant for all three data sets. On this basis, we cannot strongly recommend these metrics for quantifying knowledge structures from think-aloud data, although the prevalence of edge density in the literature (9 articles in our review) might support its use by researchers applying NA to think-aloud data.

Several aspects of the findings appear to be driven by the specific texts that we gave these specific students. First, the importance of betweenness centrality for specific nodes appears to be driven simultaneously by the conceptual importance of these concepts in all three data sets, but also in Study 1 because these specific terms links the 3 textbook subsections. For example, in the first passage a splinter is shown bringing bacteria into the body and via a sequence of signals the phagocytes produce cytokines, in the second passage bacteria active the CD14 receptor on a T cell and through a long chain of reactions cytokines are produced, and in the third passage bacteria attach to the variable segments on the short arms of an antibody molecule(s). Likewise, the Antibody node in our master list of propositions from Study 2 AH had the highest PageRank score of all 9 focal nodes across the 3 data sets when we analyzed the master list; this suggests that nodes in the original text that are connected in more triads could possibly be associated with participant PageRank scores being related to comprehension, compared to nodes in the original text that are involved in few triads. Density also may depend on the nature of the content being learned and how it is represented in the texts/stimuli, since long causal chains with few inter-connections yield lower density networks.

Second, compared to an immunologist, we knew these undergraduate students were relative novices. Therefore, we would expect a much less interconnected network (a sparser network) than we would expect from an expert, e.g., the Study 1 text uses macrophages and monocytes as examples of white blood cells, but an expert would know that these two are further connected with each other because monocytes are found in the bloodstream and later mature into macrophages in tissue. Thus, sparse networks are to be expected with novices upon one reading of a textbook excerpt and this leads to certain features of our results (e.g., low density metrics). Third, because the texts are aimed at undergraduates taking an introductory biology course, they present a level of detail that the authors believe is appropriate for novices. One could argue that the texts were impoverished (e.g., density = .054, .090, and .044 across the three master lists) and hence student knowledge networks were impoverished, but we reason that a novice would be overwhelmed by introductory text that has the level of detail that an expert would develop over decades of scientific practice. Among these participants, variability in student networks was significantly related to posttest comprehension for the majority of correlations, showing that even in this introductory text, some biology students extracted richer knowledge networks and scored better on posttest comprehension whereas others extracted very sparse networks and scored worse.

Thus, researchers applying SemNA to think-aloud data should expect the results to be somewhat specific to the learning materials, the task (after reading a passage, after a 5-week intervention unit, after a whole year of instruction?), and the expertise level of the participants.

5.1. Limitations

Our three data sets all came from undergraduate students, reading biology text about the immune system. Future work should consider whether think-aloud protocols can be fruitfully analyzed with SemNA metrics across a much wider range of topics and participants. In addition, posttests for all 3 data sets were collected immediately after completing the think-aloud on the text; in future studies the posttest should be collected after a longer delay to determine how long memory for picked-up information might last.

We used only one theory to drive our use of SemNA metrics, but other researchers analyzing during-learning data could, for example, use Expertise theory (e.g., Ericsson et al., 2018) for analyzing verbal data or eye tracking data or could use the Information Foraging model (e.g., Pirolli & Card, 1999) for analyzing eye tracking or other transition data (e.g., moving through a game or museum exhibit or a sequence of tool use in a virtual or physical environment).

For the purpose of SemNA, Expertise theory focuses on amount of knowledge (nodes) and interconnectedness of knowledge (edges), and also how knowledge is structured around key principles of a domain, such as Newton's laws for knowledge about kinematics (Chi et al., 1981). This would have implications for centrality of those conceptually key principles in expert versus novice networks. The Information Foraging (IF) model (Pirolli & Card, 1999) makes an analogy between a creature foraging for food in a natural environment and a person foraging for information in a human environment (e.g., in illustrated text, on a website, etc.). In the IF model, a person who knows what sources to forage in, who knows what cues in that source signal rich information, and who knows how to proceed through the information sources can very efficiently find information in the sense of spending less time/energy and frustration to locate needed information. For the purpose of SemNA, very efficient search (e.g., eye gaze) patterns should be associated with experienced searchers locating learning-goal relevant information, whereas less-experienced searchers may show lengthy but systematic search patterns. In sum, different theories might lead to identifying different network analysis metrics for different types of during-learning data.

5.2. Potential future applications

Here, we provided 3 examples of applying Network Analysis to think-aloud data, but our literature review includes examples of brief interview, concept map, dyadic discourse, and open-ended written response data. The studies we reviewed reported on data collected on varied learning materials, over varying time spans (from 1 hour to 1 year), with participants at differing levels of expertise. In addition to these 4 types of data, other researchers might want to use content of longer interviews, one person's social media contributions, or LMS chat over a time period to analyze that person's knowledge structures or possibly to analyze beliefs or perceptions, to analyze self-explanations, or use other kinds of knowledge data in formal or informal learning contexts. On the one hand, we believe that our literature review supports the set of metrics for this kind of individual SemNA data and that important existing metrics are not highly likely to be missing. On the other hand, we do not have the basis for saying that the metrics that were significant or non-significant for our three data sets will be significant or non-significant across all types of individual knowledge network data. This is especially true because our data in all three data sets were missing triads to a great extent, and possibly as a consequence this made the PageRank statistic non-significant. Note that we attribute this to the stimulus-task-participant constellations in our three data sets. Therefore, we recommend that analysts apply all of the metrics frequent in the literature to build a more robust knowledge base about what metrics yield valuable information for understanding learning with which types of stimuli, tasks, and participants.

Furthermore, metrics that might be useful for other kinds of during-learning data such as eye gazes, web search/navigation, movement through virtual spaces, and so on, require a different literature review to identify the frequently used metrics for those types of during-learning data. We have already found such differences in our own work in progress (Cromley & Kunze, 2024).

Future work could also apply automated text extraction techniques (NLP, machine learning) to during-learning data to try to reduce the work of hand extracting lists of propositions. One such automated system is T-MOTICAR which yields only average degree from among the metrics applied here (Pirnay-Dummer, 2020). An alternative would be to use web-based learning environments where logfiles would capture note-taking or ‘type-alouds’ while learning as a representation of what is ‘picked up’. These approaches could make SemNA a much easier technique to use on individual verbal data with learning outcomes. We look forward to learning how researchers apply these techniques to and learn from additional information available from the multitude of during-learning data that educational researchers collect.

5.5 Conclusion

Ample previous research supports the idea that readers actively transforming what they read—via self-explanation, using specific strategies, answering embedded questions—is associated with better comprehension (McNamara, 2004). Our re-analyses of the content picked up during reading further supports this idea, and potentially explains why strategies are effective; those who picked up more content formed more complex during-reading knowledge structures and characteristics of knowledge structures were significantly correlated with posttest comprehension.

References

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods, 17*(3), 399.

Bodin, M. (2012). Mapping university students' epistemic framing of computational physics using network analysis. *Physical Review Special Topics-Physics Education Research, 8*(1), 010115. https://doi.org/10.1103/PhysRevSTPER.8.010115

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121-152. https://doi.org/10.1207/s15516709cog0502_2

Clariana, R.B., & Koul, R. (2008). The effects of learner prior knowledge when creating concept maps from a text passage. *International Journal of Instructional Media, 35*(2), 177-184.

Clariana, R. B., Engelmann, T., & Yu, W. (2013). Using centrality of concept maps as a measure of problem space states in computer-supported collaborative problem solving. *Educational Technology Research and Development, 61*, 423-442. https://doi.org/10.1007/s11423-013-9293-6

Clark, D. & Linn, M. C. (2003). Designing for knowledge integration: The impact of instructional time. *Journal of the Learning Sciences, 12*(4), 451-493. https://doi.org/10.1207/S15327809JLS1204_1

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587

Cromley, J. G., & Kunze, A. J. (2024). *Applying network analysis to think-aloud and eye tracking data.* Paper presented at the 2024 Annual Meeting of the American Educational Research Association, April 11-14, 2024, Philadelphia, PA.

Cromley, J. G., Kunze, A., & Dane, A. (2021). Multi-text multi-modal reading processes and comprehension. *Learning and Instruction, 71* [February]. https://doi.org/10.1016/j.learninstruc.2020.101413

Cromley, J. G., Ma, S., Van Boekel, M., & Dane, N. (2020). Pickup of causal language and inference during and after reading illustrated text. *Reading Psychology*, *41*(3), 157-182. https://doi.org/10.1080/02702711.2020.1768974

Daems, O., Erkens, M., Malzahn, N., & Hoppe, H. U. (2014). Using content analysis and domain ontologies to check learners' understanding of science concepts. *Journal of Computers in Education, 1*, 113-131. https://doi.org/10.1007/s40692-014-0013-y

Dauer, J. T., Bergan-Roller, H. E., King, G. P., Kjose, M., Galt, N. J., & Helikar, T. (2019). Changes in students' mental models from computational modeling of gene regulatory networks. *International Journal of STEM Education, 6*(1), 1-12. https://doi.org/10.1186/s40594-019-0193-0

Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams, A. M. (Eds.). (2018). *The Cambridge handbook of expertise and expert performance.* Cambridge University Press. https://doi.org/10.1017/9781316480748

Figueres-Esteban, M., Hughes, P., & Van Gulijk, C. (2016). Visual analytics for text-based railway incident reports. *Safety Science, 89*, 72-76. https://doi.org/10.1016/j.ssci.2016.05.009

Fishman, N., & Davis, N. T. (2022). Change we can believe in: Structural and content dynamics within belief networks. *American Journal of Political Science, 66*(3), 648-663. https://doi.org/10.1111/ajps.12626

Freedman, H., Young, N., Schaefer, D., Song, Q., van der Hoek, A., & Tomlinson, B. (2024). Construction and analysis of collaborative educational networks based on student concept maps. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1-22. https://dl.acm.org/doi/pdf/10.1145/3637313

Frerichs, L., Young, T. L., Dave, G., Stith, D., Corbie-Smith, G., & Lich, K. H. (2018). Mind maps and network analysis to evaluate conceptualization of complex issues: A case example evaluating systems science workshops for childhood obesity prevention. *Evaluation and Program Planning, 68*, 135-147. https://doi.org/10.1016/j.evalprogplan.2018.03.003

Giabbanelli, P. J., & Tawfik, A. A. (2021). How perspectives of a system change based on exposure to positive or negative evidence. *Systems, 9*(2), 23. https://doi.org/10.3390/systems9020023

Gobbo, C., & Chi, M. (1986). How knowledge is structured and used by expert and novice children. *Cognitive Development, 1*(3), 221-237. https://doi.org/10.1016/S0885-2014(86)80002-8

Gogus, A., Koszalka, T. A., & Spector, J. M. (2009). Assessing conceptual representations of ill-defined problems. *Technology, Instruction, Cognition and Learning, 7*(1), 1-20.

Grebitus, C., & Bruhn, M. (2008). Analyzing semantic networks of pork quality by means of concept mapping. *Food Quality and Preference*, *19*(1), 86-96. https://doi.org/10.1016/j.foodqual.2007.07.007

Guerra, J., Huang, Y., Hosseini, R., & Brusilovsky, P. (2015). Graph analysis of student model networks. In *CEUR Workshop Proceedings* (Vol. 1446). University of Pittsburgh.

Hähnlein, I. S., & Pirnay-Dummer, P. (2024). Promoting pre-service teachers' knowledge integration from multiple text sources across domains with instructional prompts. *Educational Technology Research and Development*, 72, 2159–2185. https://doi.org/10.1007/s11423-024-10363-z

Haiyue, J., & Khoon Yoong, W. (2010). A network analysis of concept maps of triangle concepts. In L. Sparrow, B. Kissane, & C. Hurst (Eds.), *Shaping the future of mathematics education: Proceedings of the 33rd annual conference of the Mathematics Education Research Group of Australasia*. Fremantle: MERGA.

Hecking, T., Dimitrova, V., Mitrovic, A., & Hoppe, U. (2017). Using network-text analysis to characterise learner engagement in active video watching. In *ICCE 2017 Main Conference Proceedings* (pp. 326-335). Asia-Pacific Society for Computers in Education.

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science, 28*(1), 127-138. https://doi.org/10.1016/S0364-0213(03)00065-X

Hoppe, H. U., Engler, J., & Weinbrenner, S. (2012). The impact of structural characteristics of concept maps on automatic quality measurement. In van Aalst, J., Thompson, K., Jacobson, M. J., & Reimann, P. (Eds.), *The Future of Learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012) – Volume 1.* International Society of the Learning Sciences. https://doi.dx.org/10.22318/icls2012.1.291

Ifenthaler, D., Masduki, I., & Seel, N. M. (2011). The mystery of cognitive structure and how we can detect it: tracking the development of cognitive structures over time. *Instructional Science, 39*, 41-61. https://doi.org/10.1007/s11251-009-9097-6

Irby, S. M., Phu, A. L., Borda, E. J., Haskell, T. R., Steed, N., & Meyer, Z. (2016). Use of a card sort task to assess students' ability to coordinate three levels of representation in chemistry. *Chemistry Education Research and Practice*, *17*(2), 337-352. https://doi.org/10.1039/C5RP00150A

Jamieson, P. (2012, October). Using modern graph analysis techniques on mind maps to help quantify learning. In *2012 Frontiers in Education Conference Proceedings* (pp. 1-6). IEEE. https://doi.org/10.1109/FIE.2012.6462222

Kang, G. J., Ewing-Nelson, S. R., Mackey, L., Schlitt, J. T., Marathe, A., Abbas, K. M., & Swarup, S. (2017). Semantic network analysis of vaccine sentiment in online social media. *Vaccine*, *35*(29), 3621-3638. https://doi.org/10.1016/j.vaccine.2017.05.052

Kapuza, A., Koponen, I. T., & Tyumeneva, Y. (2020). The network approach to assess the structure of knowledge: Storage, distribution and retrieval as three measures in analysing concept maps. *British Journal of Educational Technology, 51*(6), 2574-2591. https://doi.org/10.1111/bjet.12938

Kim, K. (2024). How the relationship between verbal and nonverbal text signals influences second language expository text comprehension: knowledge structure analysis. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-024-10405-6

Kim, M. K., & McCarthy, K. S. (2021a). Improving summary writing through formative feedback in a technology-enhanced learning environment. *Journal of Computer Assisted Learning, 37*(3), 684-704. https://doi.org/10.1111/jcal.12516

Kim, M. K., & McCarthy, K. S. (2021b). Using graph centrality as a global index to assess students' mental model structure development during summary writing. *Educational Technology Research and Development, 69*, 971-1002. https://doi.org/10.1007/s11423-021-09942-1

Kohn, K. P., Underwood, S. M., & Cooper, M. M. (2018). Connecting structure–property and structure–function relationships across the disciplines of chemistry and biology:

Exploring student perceptions. *CBE—Life Sciences Education, 17*(2), ar33. https://doi.org/10.1187/cbe.18-01-0004

Kubsch, M., Nordine, J., Neumann, K., Fortus, D., & Krajcik, J. (2019). Probing the relation between students' integrated knowledge and knowledge-in-use about energy using network analysis. *Eurasia Journal of Mathematics, Science and Technology Education, 15*(8), em1728. https://doi.org/10.29333/ejmste/104404

Laukkanen, M. (2023). Understanding the contents and development of nascent entrepreneurs' belief systems. *International Entrepreneurship and Management Journal, 19*, 1289-1312. https://doi.org/10.1007/s11365-023-00862-5

Lehmann, T., Pirnay-Dummer, P., & Schmidt-Borcherding, F. (2020). Fostering integrated mental models of different professional knowledge domains: Instructional approaches and model-based analyses. *Educational Technology Research and Development, 68*, 905-927. https://doi.org/10.1007/s11423-019-09704-0

Linn, M. (2006). The knowledge integration perspective on learning and instruction. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 243–264). Cambridge University Press. https://doi.org/10.1017/CBO9780511816833

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*(1), 1-30. https://doi.org/10.1207/s15326950dp3801_1

Morrison, D., Bedinger, M., Beevers, L., & McClymont, K. (2022). Exploring the raison d'etre behind metric selection in network analysis: A systematic review. *Applied Network Science, 7*(1), 1-25. https://doi.org/10.1007/s41109-022-00476-w

Pirnay-Dummer, P. (2020). Knowledge and structure to teach: A model-based computer-linguistic approach to track, visualize, compare and cluster knowledge and knowledge integration in pre-service teachers. In T. Lehmann (Ed.), *International perspectives on knowledge integration* (pp. 133-154). Brill. https://doi.org/10.1163/9789004429499_007

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*(4), 643-675. https://doi.org/10.1037/0033-295X.106.4.643

Podschuweit, S., & Bernholt, S. (2020). Investigating network coherence to assess students' conceptual understanding of energy. *Education Sciences, 10*(4), 103. https://doi.org/10.3390/educsci10040103

Pokorny, J. J., Norman, A., Zanesco, A. P., Bauer-Wu, S., Sahdra, B. K., & Saron, C. D. (2018). Network analysis for the visualization and analysis of qualitative data. *Psychological Methods, 23*(1), 169-183. https://doi.org/10.1037/met0000129

Sadava, D. E., Hillis, D. M., Heller, H. C., & Bernbaum, M. (2012). *Life: The science of biology (10th ed.).* W. H. Freeman.

Salmon, P. M., Lenné, M. G., Young, K. L., & Walker, G. H. (2013). An on-road network analysis-based approach to studying driver situation awareness at rail level crossings. *Accident Analysis & Prevention, 58*, 195-205. https://doi.org/10.1016/j.aap.2012.09.012

Schwendimann, B. A. (2014). Making sense of knowledge integration maps. In D. Ifenthaler & R. Hanewald (Eds.), *Digital knowledge maps in education: Technology enhanced support for teachers and learners.* Springer. https://doi.org/10.1007/978-1-4614-3178-7_2

Schwendimann, B. A., & Linn, M. C. (2016). Comparing two forms of concept map critique activities to facilitate knowledge integration processes in evolution education. *Journal of Research in Science Teaching, 53*(1), 70-94. https://doi.org/10.1002/tea.21244

Scott, J. (2017). *Social network analysis (4th Ed.)*: Sage. https://doi.org/10.4135/9781529716597

Siew, C. S. (2019). Using network science to analyze concept maps of psychology undergraduates. *Applied Cognitive Psychology, 33*(4), 662-668. https://doi.org/10.1002/acp.3484

Siew, C. (2022). Investigating cognitive network models of learners' knowledge representations. *Journal of Learning Analytics*, *9*(1), 120-129. https://doi.org/10.18608/jla.2022.7671

Smith, R. A., & Parrott, R. L. (2012). Mental representations of HPV in Appalachia: Gender, semantic network analysis, and knowledge gaps. *Journal of Health Psychology, 17*(6), 917-928. https://doi.org/10.1177/1359105311428534

Wagner, S., Kok, K., & Priemer, B. (2020). Measuring characteristics of explanations with element maps. *Education Sciences, 10*(2), 36. https://doi.org/10.3390/educsci10020036

Wagner, S., & Priemer, B. (2023). Assessing the quality of scientific explanations with networks. *International Journal of Science Education,* 1-25. https://doi.org/10.1080/09500693.2023.2172326

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440-442. https://doi.org/10.1038/30918

Wei, W., & Yue, K. B. (2016). Using concept maps to assess students' meaningful learning in is curriculum. In *Proceedings of the EDSIG Conference, 2*(4046).

Wei, Z., Zhang, Y., Clariana, R. B., & Chen, X. (2024). The effects of reading prompts and of post-reading generative learning tasks on multiple document integration: evidence from concept network analysis. *Educational Technology Research and Development*, *72*(2), 661-685. https://doi.org/10.1007/s11423-023-10326-w

Wooldridge, A. R., Carayon, P., Shaffer, D. W., & Eagan, B. (2018). Quantifying the qualitative with epistemic network analysis: a human factors case study of task-allocation communication in a primary care team. *IISE Transactions on Healthcare Systems Engineering*, *8*(1), 72-82. https://doi.org/10.1080/24725579.2017.1418769

Yang, Z., Zhang, Y., Wang, K., & Zhu, M. (2018). Connecting with fundamental mathematical knowledge directly: the organizational features of good mathematical cognitive structure. *Frontiers in Psychology, 9,* 2267. https://doi.org/10.3389/fpsyg.2018.02267

Yang, Z., Zhu, M., Qu, Z., & Zhang, Y. (2017). Research on organization of mathematics knowledge in good mathematical cognitive structure. *EURASIA Journal of Mathematics, Science and Technology Education, 14*(1), 291-302. https://doi.org/10.12973/ejmste/78242

Zhou, Q., Wang, T., & Zheng, Q. (2015). Probing high school students' cognitive structures and key areas of learning difficulties on ethanoic acid using the flow map method. *Chemistry Education Research and Practice, 16*(3), 589-602. https://doi.org/10.1039/C5RP00059A

Supplementary Tables for *Three Applications of Semantic Network Analysis to Individual Student Think-Aloud Data*

Supplementary Table 1

Studies located in the literature review of semantic network analysis applied to learning-relevant outcomes

Cite	Notes	Number of nodes	Number of edges	Average degree = Edges/Nodes	Degree centrality	Betweenness centrality	PageRank Centrality	Diameter	Edge density Aka Network density, Density	Mean distance	LCC proportion (proportion of nodes in largest connected component)	Graph centralization	Number of clusters
1. Salmon et al., 2013	TA 15 experienced and 9 novice drivers Co-occurrence				Key concepts (a la sociometric status) differ between experts and novices			smaller is better *	larger is better *				
2. Kubsch et al., 2019	Interviews 30 middle school science students Co-occurring ideas within one explanation				Total degree for one node considered to be critical for understanding, higher is better *				larger is better *				
3. Podschuweit & Bernholt, 2020	Dyad discourse 28 high school students taking chemistry	Ns											
4. Bodin, 2012	Interviews before after for 6 physics students	More is better*	More is better*				For specific sets of nodes, higher is better*						
5. Siew, 2019	Concept maps of 101 Psych undergrads Drew immediately after learning in class	Ns								Larger is better*	Ns		
6. Clariana et al., 2013	Concept maps of 40 undergraduates Drew after one of two collaborative learning conditions											Ns	
7. Kapuza et al., 2020 Presents 3 novel network metrics	Concept maps of experts (gathered once) and of students gathered early in a year long course and 6 months later	More is better *		Average degree higher is better *	Hubs – outdegree higher is better * Authority – indegree lower is better * Authorities/Hubs, lower is better*								
8. Schwendimann 2014	Concept maps of students over the course of a school year, weighted by KI level of concepts				Outdegree higher is better * Indegree higher is better *				Higher is better				
9. Haiyue & Yoong, 2010	Concept maps of students on the topic of triangles				Group_Outdegree higher is better *								

Cite	Notes	Number of nodes	Number of edges	Average degree = Edges/Nodes	Degree centrality	Betweenness centrality	PageRank Centrality	Diameter	Edge density Aka Network density, Density	Mean distance	LCC proportion (proportion of nodes in largest connected component)	Graph centralization	Number of clusters
					Group_Indegree higher is better *								
10. Hoppe et al 2012	Concept maps on climate change, scored relative to an expert map	More is better *	More is better *	higher is better*									
11. Wei and Yue, 2016	Concept maps made in 5 Information Systems courses	More is better*	More is better*	Edges/nodes ratio mixed results	Mixed			Longer is better *					
12. Wagner and Priemer, 2023	Concept maps of physics content made by 9 experts and by 56 students	More is better*				Higher is better, for critical concepts*		Ns Diameter/size ratio, smaller is better*					Fewer is better*
13. Pokorny et al 2018	Interviews with meditation retreat participant, Purpose in Life questionnaire						Higher is worse * (recalculated from their data by ranking)						
14. Daems et al., 2014	Student-given verbal explanations as part of workshops, Expert ontologies				Three main topics of workshop*	Terms that link the 3 topics*							
15. Hecking et al., 2017	Student annotations on STEM videos, 3 clusters sorted by posttest 'soft skills' knowledge				Which words have highest centrality differs between groups*								
16. Smith & Parrott, 2013	Open-ended student knowledge about Human Papilloma Virus	Vaccinated women > Men > Unvaccinated women			Normalized in- and out-degree centrality for cancer Vaccinated women > Men > Unvaccinated women								
17. Gobbo & Chi 1986	5 more-knowledgeable and 5 less-knowledgeable 7 year old boys (sy what you know about dinosaurs)	Ns	Exp>Nov		Exp>Nov								

Cite	Notes	Number of nodes	Number of edges	Average degree = Edges/Nodes	Degree centrality	Betweenness centrality	PageRank Centrality	Diameter	Edge density Aka Network density, Density	Mean distance	LCC proportion (proportion of nodes in largest connected component)	Graph centralization	Number of clusters
18. Gogus et al., 2009	Expert and novice concept maps about two science topics	Exp>Nov (both topics)	Exp>Nov (both topics)		Exp>Nov –not formally degree centrality-- (both topics)								
19. Lehmann et al., 2020	Student essays in a 2 x 2 experimental design			Prompted>Unprompted*									
20. Guerra et al., 2005	8913 question/problem attempts by 83 students over 3 semesters of Java programming.		When weighted by success at question *	Ns				Ns	Ns	Higher is better*			
21. Jamieson, 2012	Mind maps (nodes and unlabeled edges) of								Ns				
22. Kim & McCarthy, 2021 (Improving summary writing)	Change over time in (revised) essays of graduate instructional design students	Increases*	Increases*	Increases*				Ns	Increases*	ns			
23. Ifenthaler et al, 2011	25 graduate students taking a research methods course, concept maps 5 times over semester	Ns with grade Increases* over semester		Higher is better* No change over semester				Ns with grade Increases * over semester	Ns with grade				Ns with grade Increases* over semester
24. Dauer et al., 2019	289 life science majors making conceptual models (concept maps with causal arrows) before and after an instructional video on content	Increases after instruction*	Increases after instruction*		For key nodes, increases and decreases after instruction (no stat test)								

Cite	Notes	Number of nodes	Number of edges	Average degree = Edges/Nodes	Degree centrality	Betweenness centrality	PageRank Centrality	Diameter	Edge density Aka Network density, Density	Mean distance	LCC proportion (proportion of nodes in largest connected component)	Graph centralization	Number of clusters
25. Yang, Zhang et al., 2018	403 undergraduate calculus students given a list of concepts, connect and weight the strength of connections among the 82	Not reported—from among 82 fixed/provided nodes	Best-achieving students > Middle students = Worst students *(stat test)	Best-achieving students > Middle students > Worst students (no stat test)	For 4 high-centrality nodes noted by all groups, Best-achieving students > Middle students > Worst students (no stat test)						number of nodes /LCC Best-achieving students > Middle students > Worst students (no stat test)		Best-achieving students < Middle students < Worst students (no stat test)
26. Frerichs et al., 2018	21 AfAm Adolescents making concept maps 3 times during a workshop on causes of obesity, weighted by number of repeated edges	Increases after instruction (p = .08)	Increases after instruction*						Increases after instruction (ns from .36 to .44)				
27. Zhou et al., 2015	30 HS (age 15) students interviewed after a chemistry lesson, followed by exam	Ns	More is better*										
28. Giabbanelli & Tawfik, 2021	28 CS students presented with a positive or negative case in PBL	Increased (no stat test)	Increased (no stat test)	Mostly increased (no stat test), but mixed				Increased (no stat test)		Mostly increased (no stat test), but mixed			
29. Yang, Zhu et al., 2018	206 high school (16 YO) trigonometry students given a list of concepts, connect and weight the strength of connections among the 44			Best-achieving students > Middle students = Worst students* (stat test)					Best-achieving students > Middle students > Worst students (no stat test)		number of nodes /LCC Best-achieving students > Middle students > Worst students (no stat test)		Best-achieving students < Middle students < Worst students (no stat test)
30. Wagner, Kok, & Priemer, 2020	(possibly a subset of Wagner & Priemer, 2023) 3 experts and 19 students providing	Experts>Novices				Experts>Novices (experts have a few high-betweenness nodes, novices		Experts>Novices Ratio D/numNodes also Experts>Novices					

Cite	Notes	Number of nodes	Number of edges	Average degree = Edges/Nodes	Degree centrality	Betweenness centrality	PageRank Centrality	Diameter	Edge density Aka Network density, Density	Mean distance	LCC proportion (proportion of nodes in largest connected component)	Graph centralization	Number of clusters
	written explanations of refraction in water					have no nodes high on betweenness)							
31. Laukkanen, 2023	7 new entrepreneurs interviewed before and after launching a startup (learning about running a new business)	Increases	Increases		Increases								
32. Kim & McCarthy, 2021 (Using graph centrality…	32 graduate students in an online course who revised their writing	No change	No change	No change				No change	Increases over time	No change		Experts < Novices, decrease over time	

S Supplementary Table 2

Lists of propositions for all three data sets

Study 1 proposition list

Source	Target	verb
1. Ab	Antigen	neutralize
2. Ab	Antigen	binds
3. Ab	Antigen	binds
4. Ab	HeavyChains	Have
5. Ab	Immunoglobulins	Are
6. Ab	LargeComplexes	Form
7. Ab	LightChains	Have
8. Ab	PolypeptideChains	Have
9. Ab	TwoAntigenMolecules	Binds
10. AntBindSites	Antigen	match shape
11. AntBindSites	Ab	are on outer end of
12. AntBindSites	AntBindSites	Are
13. Antigen	Bacterium	is part of
14. Antigen	Pollen	is part of
15. Antigen	Virus	is part of
16. Bacterium	CD14	binds to
17. Bcell	Ab	Makes
18. BloodVessel	Complement	Release
19. BloodVessel	Heat	Cause
20. BloodVessel	Phagocytes	Admits
21. BloodVessel	Redness	Cause
22. Body	DefenseProteins	Makes
23. Brain	Fever	Produce
24. CD14	CellMembrane	sits outside
25. CD14	TollLikeReceptor	Activates
26. CD14	TollLikeReceptor	Binds
27. CD14	WBCs	is found on
28. Complement	Phagocytes	Attracts
29. Complement	Tissue	Enters
30. ConstantRegion	DestinationAndFunction	Determine
31. ConstantRegion	SameAASequence	Has
32. Cytokine	Brain	Signals
33. Cytokine	Fever	Produce
34. Cytokine	ImmmunSys	Activates
35. Cytokine	InfectedCell	Kills
36. CytoTcell	Antigen	Engulfs
37. DamagedTissue	MastCells	Attracts

Source	Target	verb
38. DefenseProteins	Complement	Includes
39. DefenseProteins	Cytokine	Includes
40. DefenseProteins	Interferon	Includes
41. DefenseProteins	Phagocytosis	Regulate
42. Epitope	Epitope	Is
43. Epitope	ForeignProtein	Is
44. Epitope	VariableRegion	Binds
45. Fever	ImmunSys	Increases
46. Fever	Lymphocytes	Increases
47. Fever	Pathogens	Inhibits
48. FungalCellWall	CD14	binds to
49. GrowthFactor	Tissue	Signals
50. HeavyChains	Ab	is most of
51. HeavyChains	DisulfideBond	connected by
52. HeavyChains	Identical	Are
53. HeavyChains	Inside	Located
54. HelperT	Antigen	Signals
55. HelperT	Bcell	Signals
56. HelperT	CytoTcell	Signals
57. Histamine	AllergicRx	Causes
58. Histamine	BloodVessel	Dilates
59. Histamine	BloodVessel	Enters
60. Histamine	BloodVessel	Leakifies
61. Histamine	Inflammation	Causes
62. HIV	HelperT	Damages
63. ImmuneResponse	Non-specific	Has
64. ImmuneResponse	Specific	Has
65. ImmunSys	Ab	is part of
66. ImmunSys	Bcell	is part of
67. ImmunSys	CellSignaling	Signals
68. ImmunSys	CytoTcell	is part of
69. ImmunSys	HelperT	is part of
70. Infection	Inflammation	Causes
71. Inflammation	Damage	Isolates
72. Inflammation	MastCells	Attracts
73. Inflammation	Molecules	Recruits
74. Inflammation	Pain	Cause
75. Inflammation	Phagocytes	Recruits
76. Inflammation	Spread	Stops
77. Injury	Inflammation	Causes
78. LargeComplexes	EasyTargets	Are
79. LightChains	Identical	Are

Source	Target	verb
80. LightChains	Outside	Located
81. Lymphocyte	ImmunSys	part of
82. MastCells	Cytokine	Release
83. MastCells	Damage	go to
84. MastCells	Early	Respond
85. MastCells	Histamine	Release
86. MastCells	Histamine	stop release
87. MastCells	Phagocytes	Signals
88. MastCells	Skin	adhere to
89. MoreTissue	Scab	Forms
90. NfkB	NuclearFactorKappaBCells	Is
91. NfkB	Nucleus	Enters
92. NfkB	Nucleus	straightened enters
93. NfkB	Promoters	binds to
94. Pathogen	Body	Invades
95. Pathogen	ImmunSys	Signals
96. Pathogen	MolecularChange	binding causes
97. Phagocytes	Cytokine	Produce
98. Phagocytes	Damage	go to
99. Phagocytes	DeadCells	Engulf
100. Phagocytes	Healing	do most
101. Phagocytes	ImmunSys	part of
102. Phagocytes	LargeComplexes	Destroy
103. Phagocytes	LargeComplexes	Ingest
104. Phagocytes	Pathogens	Engulf
105. Phagocytes	Tissue	Enters
106. Plasma	Swelling	Causes
107. Plasma	Tissue	Enters
108. PolypeptideChains	ConstantRegion	Have
109. PolypeptideChains	VariableRegion	Have
110. Promoters	DefenseProteins	Makes
111. Promoters	Transcription	Start
112. Prostoglandins	BloodVessel	Dilate
113. Prostoglandins	Nerves	increase sensitivity of
114. ProteinKinaseCascade	CellMembrane	happens inside
115. ProteinKinaseCascade	FortyGenesTranscribed	Causes
116. ProteinKinaseCascade	NfkB	Straightens
117. Splinter	BloodVessel	Injures
118. Splinter	DamagedCells	Has
119. Splinter	Pathogens	Has
120. Tcell	Pathogen	Binds

Source	Target	verb
121. Tissue	MoreTissue	Divides
122. TNF	Cytokine	is a
123. TNF	ImmmunSys	Activates
124. TNF	InfectedCell	Kills
125. TollLikeReceptor	CellMembrane	passes through
126. TollLikeReceptor	Development	involved in
127. TollLikeReceptor	ImmuneResponse	involved in
128. TollLikeReceptor	ProteinKinaseCascade	Starts
129. Transcription	DefenseProteins	Makes
130. VariableRegion	OnOuterEndOfAb	Are
131. VariableRegion	Specificity	responsible for
132. WBCs	Macrophages	Includes
133. WBCs	Monocytes	Includes

Study 2 AH text set proposition list

Source	Target	Verb	Passage
1. Ab	Ab	is	P1
2. Ab	Ab	links	P1
3. Ab	IgE	include	P1
4. Ab	MastCell	binds	P1
5. Ab	Receptors	binds	P1
6. Allergens	Ab	causes	P1
7. Allergens	Allergens	is	P1
8. AllergicRx	AllergicRx	is	P1
9. Antigen	Ab	binds	P1
10. Antigen	Antigen	is	P1
11. Antihistamines	Histamine	block	P1
12. Bcell	BCell	is	P1
13. Histamine	AllergicRx	causes	P1
14. Histamine	Histamine	is	P1
15. IgE	IgE	is	P1
16. IgE	Immunoglobulin	istype	P1
17. MastCell	Histamine	makes	P1
18. MastCell	Histamine	release	P1
19. MastCell	MastCell	is	P1
20. MastCell	Receptors	has	P1
21. MastCell	Vesicles	has	P1
22. PlasmaCells	Ab	make	P1
23. Pollen	AllergicRx	causes	P1
24. Vesicles	Histamine	release	P1

Source	Target	Verb	Passage
25. Ab	Ab	links	P2
26. Ab	Antigens	attackspecific	P2
27. Ab	Basophils	binds	P2
28. Ab	BCell	binds	P2
29. Ab	ConstantEnd	has	P2
30. Ab	MastCell	binds	P2
31. AllergicRx	Death	cancause	P2
32. AllergicRx	genetic	canbe	P2
33. Antigen	Ab	binds	P2
34. Antigen	Antigen	is	P2
35. Basophils	Basophils	are	P2
36. Bcell	Ab	make	P2
37. Bcell	Bcell	is	P2
38. Bcell	PlasmaCells	makes	P2
39. Genetics	AllergicRx	causes	P2
40. Histamine	DifficultyBreathing	causes	P2
41. Histamine	Histamine	is	P2
42. IgE	IgE	is	P2
43. MastCell	Histamine	makes	P2
44. MastCell	Histamine	releases	P2
45. MastCell	MastCell	is	P2
46. PlasmaCells	Ab	make	P2
47. Ab	Ab	is	P3
48. Ab	Ab	links	P3
49. Ab	MastCell	binds	P3
50. Ab	SignalTransduction	initiates	P3
51. Antigen	Antigen	is	P3
52. Antigen	Bcell	binds	P3
53. Antigen	ThCell	attaches	P3
54. APC	Antigen	presents	P3
55. APC	APC	is	P3
56. Bcell	Ab	makes	P3
57. Bcell	BCell	is	P3
58. Bcell	PlasmaCell	becomes	P3
59. Cytokine	BCell	signals	P3
60. Cytokine	Cytokine	is	P3
61. Histamine	AllergicRx	causes	P3
62. Histamine	Histamine	is	P3
63. Histamine	Signal	isa	P3
64. IL-4	Bcell	activates	P3
65. IL-4	Cytokine	isa	P3
66. IL-4	IL-4	is	P3

Source	Target	Verb	Passage
67. IL-5	IL-5	is	P3
68. MastCell	Histamine	releases	P3
69. MastCell	MastCell	is	P3
70. PlasmaCell	Ab	makes	P3
71. ThCell	Cytokine	releases	P3
72. ThCell	IL-4	releases	P3
73. ThCell	IL-5	releases	P3
74. ThCell	ThCell	is	P3
75. Ab	Ab	binds	P4
76. Ab	*Basophils*	*binds*	*P4*
77. Ab	*MastCell*	*binds*	*P4*
78. Allergen	APC	binds	P4
79. Antigen	*Ab*	*binds*	*P4*
80. APC	Antigen	neutralizes	P4
81. APC	APC	is	P4
82. APC	ThCell	signals	P4
83. Bcell	*Ab*	*make*	*P4*
84. Bcell	*Bcell*	*is*	*P4*
85. Bcell	*PlasmaCells*	*makes*	*P4*
86. Histamine	Rash	causes	P4
87. MastCell	Histamine	releases	P4
88. PlasmaCells	*Ab*	*make*	*P4*
89. ThCell	Bcell	signals	P4
90. ThCell	ThCell	is	P4

Study 2 TC text set proposition list

Source	Target	Verb	Passage
1. Ab	2Polypeptides	Madeof	P1
2. Ab	Ab	Is	P1
3. Ab	Antigens	bind_to	P1
4. Ab	Antigens	recognizesOne	P1
5. alphaChain	alphaChain	Is	P1
6. alphaChain	betaChain	Attachedto	P1
7. alphaChain	ConstantRegion	has_both	P1
8. alphaChain	exterior	is_on	P1
9. alphaChain	variableRegion	has_both	P1
10. Antigens	Proteins	Are	P1
11. Antigens	variableRegion	Bind	P1
12. Bcell	Blood	respondsTo	P1
13. betaChain	betaChain	Is	P1

Source	Target	Verb	Passage
14. ConstantRegion	ConstantRegion	Is	P1
15. ConstantRegion	Immunoglobulin	Variesbetween	P1
16. ConstantRegion	TCR	Anchors	P1
17. Glycoproteins	protein_and_sugar	Are	P1
18. hydrophobicRegion	stopWater	Means	P1
19. Immunoglobulin	Immunoglobulin	Is	P1
20. MHCs	Antigens	Display	P1
21. MHCs	Antigens	Display	P1
22. plasmaMembrane	hydrophobicRegion	Has	P1
23. Tcell	Antigens	Recognizes	P1
24. Tcell	Cell	respondsToInfected	P1
25. Tcell	self	Recognizes	P1
26. Tcell	TCR	Has	P1
27. TCR	2Polypeptides	Madeof	P1
28. TCR	alphaChain	Has	P1
29. TCR	Antigens	Bind	P1
30. TCR	Antigens	Recognizes	P1
31. TCR	BCellReceptor	Smallerthan	P1
32. TCR	betaChain	Has	P1
33. TCR	ConstantRegion	Has	P1
34. TCR	plasmaMembrane	sits_in	P1
35. TCR	Tcell	sits_outside	P1
36. TCR	TCR	Is	P1
37. TCR	variableRegion	Has	P1
38. variableRegion	Antigens	Binds	P1
39. variableRegion	TCR	Variesbetween	P1
40. variableRegion	variableRegion	Is	P1
41. AntigenBindingSite	AntigenBindingSite	Is	P2
42. AntigenBindingSite	variableRegion	Madeof	P2
43. Antigens	AntigenBindingSite	bind_at	P2
44. Antigens	Antigens	Are	P2
45. Antigens	Cell	are_onInfected	P2
46. Antigens	Pathogens	part_of	P2
47. APCs	APCs	Are	P2
48. Bcell	Bcell	Is	P2
49. Bcell	Humoral	involved_in	P2
50. Bcell	lymphNodes	found_in	P2
51. Bcell	spleen	found_in	P2
52. ConstantRegion	ConstantRegion	Is	P2
53. ConstantRegion	TCR	sits_at_bottom	P2
54. DisulfideBridge	ConstantRegion	Connects	P2
55. DisulfideBridge	DisulfideBridge	Is	P2

Source	Target	Verb	Passage
56. immatureBcell	self	Attacks	P2
57. immatureBcell	self	binds_strongly	P2
58. immatureBcell	self	binds_weakly	P2
59. ImmuneSystem	Diseases	trains_on	P2
60. IntracellularDomain	IntracellularDomain	Is	P2
61. MHCs	Antigens	Display	P2
62. MHCs	Cell	are_inside	P2
63. MHCs	MajorHistcompatibilityComplex	Are	P2
64. MHCs	*MHCs*	*Is*	*P2*
65. Pathogens	Antigens	Madeof	P2
66. Pathogens	Cell	Infect	P2
67. Tcell	Antigens	Recognizes	P2
68. Tcell	Cell-mediated	involved_in	P2
69. Tcell	immuneResponse	Initiates	P2
70. Tcell	MHCs	Has	P2
71. Tcell	Pathogens	Recognizes	P2
72. Tcell	Tcell	Is	P2
73. TCR	alphaChain	Has	P2
74. TCR	betaChain	Has	P2
75. TCR	Tcell	sits_outside	P2
76. TCR	ConstantRegion	Has	P2
77. TCR	plasmaMembrane	sits_in	P2
78. TCR	variableRegion	Has	P2
79. transmembraneRegion	transmembraneRegion	Is	P2
80. variableRegion	TCR	sits_at_top	P2
81. variableRegion	variableRegion	Is	P2
82. Alphabeta	TCR	form_of	P3
83. alphaChain	alphaChain	Is	P3
84. AntigenBindingSite	*variableRegion*	*Madeof*	*P3*
85. Antigens	epitope	Is	P3
86. Antigens	MHCs	binds_to	P3
87. betaChain	betaChain	Is	P3
88. C_terminus	TCR	sits_at_bottom	P3
89. Calpha	Calpha	Is	P3
90. carbohydrateGroup	carbohydrateGroup	Is	P3
91. carbohydrateGroup	variability	Has	P3
92. Cbeta	Cbeta	Is	P3
93. ConstantRegion	constantRegion	Is	P3
94. DisulfideBridge	*ConstantRegion*	*Connects*	*P3*
95. DisulfideBridge	*DisulfideBridge*	*Is*	*P3*
96. Epitope	epitope	Is	P3
97. Gammadelta	mucosal	found_on	P3

Source	Target	Verb	Passage
98. Heterodimer	heterodimer	Is	P3
99. MHCs	MHCs	Is	P3
100. N_terminus	TCR	sits_at_top	P3
101. TCR	alphaChain	Has	P3
102. TCR	AntigenBindingSite	Has	P3
103. TCR	Antigens	binds_to	P3
104. TCR	Antigens	Recognizes	P3
105. TCR	betaChain	Has	P3
106. TCR	C_terminus	Has	P3
107. TCR	Cell	sits_outside	P3
108. TCR	ConstantRegion	Has	P3
109. TCR	N_terminus	Has	P3
110. TCR	variableRegion	Has	P3
111. TCR	variableRegion	has_at_top	P3
112. transmembraneRegion	*transmembraneRegion*	*Is*	*P3*
113. Valpha	Valpha	Is	P3
114. Vbeta	Vbeta	Is	P3
115. variableRegion	Antigens	Determines	P3
116. alphaChain	betaChain	Attachedto	P4
117. AntigenBindingSite	loops	formed_of	P4
118. AntigenBindingSite	TCR	at_tip_of	P4
119. AntigenBindingSite	variablility	Has	P4
120. Diseases	Cell	causeInfected	P4
121. DisulfideBridge	*ConstantRegion*	*Connects*	*P4*
122. hydrophobicRegion	hydrophobicRegion	Is	P4
123. ImmuneSystem	Cell	killsInfected	P4
124. ImmuneSystem	Diseases	Finds	P4
125. ImmuneSystem	self	doesn't_attack	P4
126. Lymphocytes	lymphocytes	Are	P4
127. MHCs	Antigens	Display	P4
128. MHCs	MHCs	Is	P4
129. Tcell	Antigens	recognizesMultiple	P4
130. TCR	Ab	half_size_of	P4
131. TCR	alphaChain	Has	P4
132. TCR	Antigens	binds_to	P4
133. TCR	betaChain	Has	P4
134. TCR	ConstantRegion	Has	P4
135. TCR	plasmaMembrane	sits_in	P4
136. TCR	variableRegion	Has	P4
137. variableRegion	Pathogens	protects_against	P4
138. variableRegion	variableRegion	Are	P4

Supplementary Table 3

Network analysis metrics applied to verbal data, definitions, and igraph *code*

Metric	*definition in Semantic Network Analysis*	*code in* igraph
Node level		
nodes	number of facts, nouns, noun phrases that exist in a person's knowledge network	gorder(g)
Edges	connections between nodes in whole knowledge network, which could be verbs in phrases, links in concept maps	gsize(g)
average degree	number of edges/number of nodes in whole knowledge network	Calculate gsize(g)/gorder(g)
degree centrality	a count of all edges for focal node in a knowledge network; in-degree if pointing into that node and outdegree if pointing out from that node	degree(g, v=V(g)["Antigens"], mode = "in", loops=FALSE) degree(g, v=V(g)["Antigens"], mode = "out", loops=FALSE)
betweenness centrality	a count of the number of paths between other nodes where that path passes through a focal node in a knowledge network	betweenness(g, v=V(g)["Antigens"])

Metric	*definition in Semantic Network Analysis*	*code in* igraph
PageRank centrality	a measure of how influential the nodes are that are connected to a focal node in a knowledge network	page_rank(g, v=V(g)["Antigens"], weights=NA)$vector
Network level		
Diameter	the longest path required to traverse the entire knowledge network	Unweighted: diameter(g, weights=NA) Weighted: diameter(g)
Density	how interconnected all nodes are in a knowledge network, compared to a network with all possible interconnections	edge_density(g)
mean distance	the average number of steps out per node, across all nodes in the knowledge network	Unweighted: mean_distance(g) Weighted: mean_distance(g, weights=NA)
network centralization	standard deviation of the number of edges per node, across the whole knowledge network	centralization.degree(g, loops=FALSE)$centralization

Metric	*definition in Semantic Network Analysis*	*code in* igraph
number of components	the number of internally-interconnected but externally not/little connected groupings in a knowledge network	components(g)$csize
largest connected component as percent of nodes	the number of nodes in the largest component, divided by the total number of nodes in the knowledge network	Calculate max of components(g)$csize, divide by gorder(g)

Supplementary Table 4

Descriptive statistics on all metrics across 3 data sets

Study 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	16	18	19	20	21	22	23	24	25	26
1. Mweight	--																									
2. NumNodes	.420**	--																								
3. NumEdges	.409**	.966**	--																							
4. AvgDegree	-.233*	-.701**	-.533**	--																						
5. DegreeInBacterium	.146	.378**	.348**	-.307**	--																					
6. DegreeInCytokine	.290*	.580**	.544**	-.432**	-.036	--																				
7. DegreeInPhagocyte	.394**	.483**	.412**	-.493**	.185	.267*	--																			
8. DegreeOutBacterium	.358**	.457**	.438**	-.344**	.126	.255*	.327**	--																		
9. DegreeOutCytokine	.402**	.502**	.446**	-.409**	.081	.589**	.346**	.416**	--																	
1. DegreeOutPhagocyte	.434**	.533**	.488**	-.469**	.383**	.263*	.269*	.213	.278*	--																
11. BetweennessBacterium	.264*	.498**	.429**	-.448**	.373**	.184	.427**	.699**	.358**	.349**	--															
12. BetweennessCytokine	.393**	.559**	.488**	-.495**	-.064	.700**	.443**	.394**	.826**	.354**	.364**	--														
13. BetweennessPhagocyte	.424**	.466**	.402**	-.466**	.162	.239*	.865**	.355**	.328**	.423**	.498**	.498**	--													
14. PageRankBacterium	.064	-.142	-.153	.145	.665**	-.234*	-.018	-.013	-.045	.069	.043	-.206	-.006	--												
15. PageRankCytokine	.168	.205	.176	-.154	-.112	.747**	.082	.207	.499**	.054	.071	.448**	.081	-.177	--											
16. PageRankPhagocyte	.330**	.073	.003	-.177	.208	.019	.509**	.099	.112	.409**	.275*	.128	.448**	.187	.015	--										
17. GraphDensity	-.376**	-.808**	-.921**	.232*	-.328**	-.405**	-.253*	-.434**	-.316**	-.364**	-.381**	-.329**	-.262*	.100	-.107	.065	--									
18. NumberOfClusters	.305**	.343**	.515**	.257*	.127	.079	-.067	.265*	.055	.117	.053	-.017	-.076	-.013	.025	-.114	-.736**	--								
19. LCCSize	.256*	.800**	.718**	-.753**	.477**	.402**	.474**	.379**	.339**	.582**	.539**	.436**	.511**	-.029	.135	.161	-.545**	.017	--							
2. LCCPct	-.194	.027	-.099	-.484**	.187	-.008	.220	.059	.008	.227*	.275*	.096	.271*	.019	-.060	.106	.272*	-.645**	.544**	--						
21. GraphCentralization	-.518**	-.412**	-.496**	-.035	-.070	-.226*	-.178	-.266*	-.316**	-.183	-.195	-.255*	-.183	.022	-.172	-.022	.572**	-.568**	-.111	.523**	--					
22. MDistWeight	.652**	.731**	.680**	-.575**	.301**	.395**	.514**	.434**	.425**	.441**	.491**	.566**	.583**	.008	.133	.177	-.556**	.154	.628**	.050	-.495**	--				
23. MDistUnweight	.284*	.704**	.637**	-.699**	.279*	.396**	.463**	.442**	.388**	.354**	.602**	.533**	.546**	-.117	.149	.074	-.485**	-.013	.726**	.340**	-.269*	.844**	--			
24. DiameterWeight	.483**	.740**	.694**	-.578**	.333**	.351**	.438**	.415**	.349**	.420**	.474**	.499**	.503**	-.020	.098	.083	-.580**	.176	.661**	.067	-.456**	.949**	.862**	--		
25. DiameterUnweight	.355**	.700**	.653**	-.618**	.260*	.375**	.463**	.433**	.341**	.368**	.572**	.493**	.567**	-.103	.148	.083	-.535**	.071	.701**	.242*	-.352**	.872**	.964**	.902**	--	
26. PosttestComprehension	.209	.511**	.459**	-.421**	.252*	.385**	.326**	.208	.294**	.256*	.260*	.329**	.270*	.031	.074	.165	-.342**	.031	.492**	.150	-.103	.412**	.412**	.404**	.387**	--
Mean	.001	.000	-.000	-.000	.014	.043	.069	.054	.060	.023	.074	.072	.077	.020	.050	.027	-.000	-.005	.003	-.001	.000	.000	.006	-.003	.010	-.000
Median	.016	-.065	-.016	.016	-.197	.297	-.349	-.528	-.566	.213	-.455	-.491	-.314	.000	.000	.000	.000	-.230	-.033	.000	.000	.000	.000	.065	-.147	.049
Standard Deviation	.996	.996	.997	.998	.885	.826	.723	.769	.804	.904	.788	.799	.727	.952	.881	.937	.998	.963	.987	.994	.998	.998	.984	.968	.9450	.995
25th Percentile	-.696	-.675	-.664	-.695	-.197	-.727	-.349	-.528	-.566	-1.127	-.455	-.491	-.314	-.685	-.814	-.685	-.685	-.770	-.566	-.685	-.685	-.685	-.685	-.585	-.579	-.644
75th Percentile	.685	.664	.644	.685	.770	.297	.959	.727	.454	.769	.705	.718	.196	.685	.685	.685	.685	.705	.675	.685	.685	.685	.685	.727	.547	.685

Note: Correlations above |.221| are statistically significant at $p < .05$. * indicates significance at $p < .05$; ** indicates significance at $p < .01$

Study 2 AH

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	16	18	19	20	21	22	23	24	25	26
1. Mweight	--																									
2. NumNodes	.207	--																								
3. NumEdges	.173	.850**	--																							
4. AvgDegree	.072	.350	.752**	--																						
5. DegreeInBacterium	.488**	.671**	.780**	.594**	--																					
6. DegreeInCytokine	.369*	.527**	.625**	.512**	.510**	--																				
7. DegreeInPhagocyte	.357	.408*	.645**	.660**	.655**	.436*	--																			
8. DegreeOutBacterium	.401*	.475**	.707**	.735**	.684**	.575**	.892**	--																		
9. DegreeOutCytokine	.334	.494**	.424*	.220	.482**	.600**	.335	.364*	--																	
1. DegreeOutPhagocyte	.372*	.478**	.641**	.629**	.543**	.959**	.511**	.642**	.571**	--																
11. BetweennessBacterium	.450*	.625**	.775**	.705**	.900**	.561**	.703**	.764**	.438*	.640**	--															
12. BetweennessCytokine	.317	.666**	.655**	.398*	.664**	.716**	.533**	.547**	.864**	.693**	.681**	--														
13. BetweennessPhagocyte	.418*	.707**	.757**	.554**	.813**	.689**	.633**	.647**	.658**	.733**	.891**	.885**	--													
14. PageRankBacterium	.752**	.083	.273	.416*	.576**	.132	.464**	.456*	.131	.239	.594**	.149	.390*	--												
15. PageRankCytokine	.530**	.145	.103	.076	.283	.451*	.136	.153	.239	.468**	.335	.319	.439*	.263	--											
16. PageRankPhagocyte	-.075	-.492**	-.410*	-.197	-.101	-.509**	-.210	-.279	-.230	-.476**	-.129	-.282	-.276	.180	-.150	--										
17. GraphDensity	-.060	-.582**	-.116	.448*	-.021	-.112	.217	.193	-.344	.007	.076	-.235	-.143	.324	-.060	.425*	--									
18. NumberOfClusters	-.654**	.026	-.203	-.442*	-.505**	-.383*	-.391*	-.432*	-.248	-.443*	-.633**	-.341	-.514**	-.711**	-.549**	-.153	-.463**	--								
19. LCCSize	.500**	.791**	.854**	.584**	.856**	.565**	.660**	.696**	.466**	.574**	.881**	.714**	.854**	.485**	.362*	-.283	-.112	-.512**	--							
2. LCCPct	.540**	.119	.408*	.609**	.574**	.339	.652**	.634**	.195	.404*	.692**	.388*	.545**	.666**	.417*	.022	.491**	-.861**	.669**	--						
21. GraphCentralization	.281	.012	.024	.050	.110	.153	.175	.088	.413*	.101	.136	.354	.181	.234	.048	.363*	.104	-.321	.202	.320	--					
22. MDistWeight	.447*	.548**	.649**	.582**	.758**	.561**	.575**	.578**	.514**	.650**	.877**	.735**	.888**	.480**	.507**	-.087	.056	-.649**	.778**	.611**	.178	--				
23. MDistUnweight	.446*	.558**	.700**	.647**	.767**	.545**	.689**	.659**	.467**	.633**	.894**	.726**	.873**	.496**	.489**	-.139	.120	-.646**	.851**	.732**	.220	.959**	--			
24. DiameterWeight	.460*	.600**	.652**	.524**	.751**	.493**	.550**	.573**	.334	.560**	.835**	.594**	.789**	.436*	.513**	-.147	-.001	-.578**	.784**	.561**	.107	.911**	.887**	--		
25. DiameterUnweight	.417*	.574**	.693**	.641**	.707**	.484**	.645**	.649**	.262	.564**	.879**	.549**	.778**	.493**	.437*	-.169	.094	-.603**	.820**	.678**	.115	.890**	.931**	.921**	--	
26. PosttestComprehension	.407*	.392*	.495**	.432*	.686**	.233	.413*	.407*	.173	.320	.556**	.284	.451*	.573**	.228	-.079	.102	-.343	.540**	.371*	.054	.403*	.418*	.488**	.393*	--
Mean	-.033	.004	.004	.000	.005	.022	-.007	.005	.018	.018	.012	.045	.032	.000	.011	.000	.000	.028	.001	-.042	.000	-.000	-.000	-.012	-.004	.000
Median	0.00	-0.17	0.00	0.00	0.08	0.04	-0.25	0.00	-0.13	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.04	0.00	0.00	0.00	0.00	0.00	-0.21	-0.21	0.00
Standard Deviation	0.92	0.98	0.97	0.99	0.94	0.90	0.91	0.95	0.87	0.89	0.97	0.89	0.92	1.00	0.97	1.00	0.99	0.88	0.99	0.90	0.99	1.00	1.00	0.93	0.96	1.00
25th Percentile	-0.70	-0.67	-0.67	-0.73	-0.78	-1.11	-1.04	-1.04	-1.04	-1.11	-0.73	-0.84	-1.04	-0.70	-0.70	-0.70	-0.78	-0.97	-0.73	-0.66	-0.70	-0.70	-0.70	-0.67	-0.73	-0.70
75th Percentile	1.04	0.68	0.67	0.66	0.69	0.73	0.48	0.73	0.67	0.46	0.73	0.62	0.78	0.70	0.70	0.70	0.73	0.97	0.62	0.90	0.72	0.72	0.72	0.57	0.52	0.70

Study 2 TC

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1. Mweight	--																							
2. NumNodes	.412*	--																						
3. NumEdges	.318	.879**	--																					
4. AvgDegree	.057	.178	.560**	--																				
5. DegreeInAntigen	.765**	.744**	.686**	.162	--																			
6. DegreeInTCR	.268	.445*	.537**	.404*	.468**	--																		
7. DegreeOutAntigen	.241	.549**	.470**	.073	.473**	.110	--																	
8. DegreeOutMCH	.766**	.455*	.370*	.005	.743**	.203	.358	--																
9. DegreeOutTCR	-.063	.335	.641**	.742**	.187	.451*	.015	-.098	--															
11. BetweennessBacterium	.191	.447*	.414*	.122	.394*	.103	.946**	.261	.083	--														
11. BetweennessPhagocyte	.107	.415*	.590**	.520**	.402*	.752**	.368*	.070	.613**	.474**	--													
12. PageRankBacterium	.756**	.261	.318	.197	.639**	.516**	.126	.606**	.133	.151	.365*	--												
13. PageRankCytokine	.096	-.368*	-.337	-.253	-.072	-.370*	-.176	.083	-.345	-.098	-.266	.084	--											
14. PageRankPhagocyte	.002	-.454*	-.281	.205	-.180	.455*	-.362*	-.120	.197	-.217	.361	.306	.024	--										
15. GraphDensity	-.307	-.823**	-.564**	.116	-.521**	-.184	-.405*	-.311	.071	-.287	-.126	-.036	.156	.616**	--									
16. NumberOfClusters	.114	.734**	.571**	-.022	.376*	.062	.329	.122	.066	.216	.085	-.165	-.235	-.687**	-.831**	--								
17. LCCSize	.243	.785**	.858**	.421*	.617**	.534**	.489**	.431*	.566**	.468**	.589**	.349	-.298	-.234	-.473**	.348	--							
18. LCCPct	-.255	-.296	-.030	.354	-.151	.158	-.005	-.075	.364*	.091	.302	.104	.052	.408*	.545**	-.646**	.299	--						
19. GraphCentralization	-.257	-.228	.111	.608**	-.188	.231	-.212	-.390*	.700**	-.117	.360	-.043	-.241	.464**	.602**	-.394*	.063	.534**	--					
20. MDistWeight	.397*	.490**	.588**	.488**	.540**	.483**	.631**	.369*	.388*	.715**	.759**	.473**	-.239	.107	-.216	.181	.607**	.202	.149	--				
21. MDistUnweight	.084	.395*	.540**	.508**	.348	.606**	.529**	.101	.535**	.652**	.917**	.312	-.261	.276	-.117	.111	.574**	.301	.300	.877**	--			
22. DiameterWeight	.369*	.516**	.613**	.530**	.526**	.539**	.573**	.412*	.419*	.650**	.782**	.482**	-.253	.136	-.258	.152	.660**	.224	.127	.962**	.882**	--		
23. DiameterUnweight	.168	.482**	.587**	.455*	.398*	.558**	.608**	.266	.424*	.693**	.837**	.339	-.231	.157	-.230	.182	.631**	.226	.146	.901**	.953**	.927**	--	
24. PosttestComprehension	.324	.434*	.468**	.225	.438*	.309	.467**	.478**	.317	.488**	.517**	.291	-.225	.026	-.157	.040	.591**	.266	.046	.601**	.531**	.671**	.570**	--
Mean	-.000	.001	-.003	.004	.013	.022	.063	.023	.004	.074	.033	.016	.020	.000	-.001	.003	-.001	.002	.000	.007	.033	.007	.029	.000
Median	.000	.000	.000	.000	-.126	.000	-.385	-.210	.000	-.385	-.084	.000	.000	.000	.000	.042	-.021	.000	.000	.000	.000	.000	-.210	.000
Standard Deviation	.995	.992	.984	.987	.948	.873	.729	.886	.979	.761	.918	.959	.949	.996	.994	.979	.982	.987	.995	.978	.920	.969	.902	.996
25th Percentile	-.702	-.702	-.716	-.626	-.784	-1.036	-.385	-1.036	-.635	-.385	-1.036	-.702	-.702	-.702	-.702	-.728	-.903	-.903	-.702	-.702	-1.036	-.674	-1.036	-.702
75th Percentile	.702	.663	.728	.702	.842	.903	.903	.432	.623	.728	.678	.702	.702	.702	.702	.784	.842	.702	.702	.702	.702	.573	.431	.702

[c] indicates could not be calculated

Supplementary Table 5

Results of exploratory regressions predicting posttest comprehension across 3 data sets

Study 1

Predictors entered in stepwise regression	Model 1 (*b*)	Model 2 (*b*)	Model 3 (*b*)
Number of nodes	.42*	-.62	-.65
Number of edges	—	1.13*	1.11*
PageRank Phagocytes	—	—	13.49*
R^2	.20	.30	.34
F of change in R^2	19.13*	9.70*	4.40*

Note: Metrics were entered in 6 blocks: 1) number of nodes, 2) number of edges, 3) number of clusters, 4) diameter and mean distance, 5) all centrality metrics, 6) all other metrics. * indicates statistically significant at $p < .05$.

Study 2 AH Dataset

Predictors entered in stepwise regression	Model 1 (*b*)	Model 2 (*b*)	Model 3 (*b*)
Number of nodes	.14*	-.06	-.06
Number of edges	—	.10*	< .01
Indegree Antibodies	—	—	.46*
R^2	.16	.28	.53
F of change in R^2	5.47*	4.56*	13.45*

Note: Metrics were entered in 6 blocks: 1) number of nodes, 2) number of edges, 3) number of clusters, 4) diameter and mean distance, 5) all centrality metrics, 6) all other metrics. * indicates statistically significant at $p < .05$.

Study 2 TC Dataset

Predictors entered in stepwise regression	Model 1 (*b*)	Model 2 (*b*)	Model 3 (*b*)
Number of nodes	.04*	.01	< .01
Diameter	—	.38*	.37*
Outdegree MHCs	—	—	.29*
R^2	.17	.34	.47
F of change in R^2	5.67*	7.17*	6.03*

Note: Metrics were entered in 6 blocks: 1) number of nodes, 2) number of edges, 3) number of clusters, 4) diameter and mean distance, 5) all centrality metrics, 6) all other metrics. * indicates statistically significant at $p < .05$.