Memory & Cognition
https://doi.org/10.3758/s13421-022-01314-1

Investigating the network structure of domain-specific knowledge using the semantic fluency task

Cynthia S. Q. Siew[1] · Anutra Guru[1]

Accepted: 11 April 2022

Abstract

Cognitive scientists have a long-standing interest in quantifying the structure of semantic memory. Here, we investigate whether a commonly used paradigm to study the structure of semantic memory, the semantic fluency task, as well as computational methods from network science could be leveraged to explore the underlying knowledge structures of academic disciplines such as psychology or biology. To compare the knowledge representations of individuals with relatively different levels of expertise in academic subjects, undergraduate students (i.e., *experts*) and preuniversity high school students (i.e., *novices*) completed a semantic fluency task with cue words corresponding to general semantic categories (i.e., animals, fruits) and specific academic domains (e.g., psychology, biology). Network analyses of their fluency networks found that both domain-general and domain-specific semantic networks of undergraduates were more efficiently connected and less modular than the semantic networks of high school students. Our results provide an initial proof-of-concept that the semantic fluency task could be used by educators and cognitive scientists to study the representation of more specific domains of knowledge, potentially providing new ways of quantifying the nature of expert cognitive representations.

Keywords Knowledge representation · Semantic fluency task · Semantic networks · Expertise

Cognitive scientists have a long-standing interest in quantifying the structure of semantic memory. Much research in the cognitive sciences is dedicated to empirical work that investigates the retrieval and organization of semantic memory (Anderson & Bower, 1980; Reder et al., 2009; Tulving, 1972), as well as methodological and computational work that focusses on quantifying or estimating memory representations (De Deyne & Storms, 2008; Jones & Mewhort, 2007; Landauer, 2007). In this paper, we investigate whether the combination of behavioral methods from cognitive psychology and computational methods from network science could be leveraged to quantify underlying knowledge representations of academic disciplines, such as psychology or biology.

Understanding the nature of domain-specific (or subject-specific) knowledge structures is an important question within the educational sciences, where it is commonly recognized that students' domain-specific knowledge should consist of a collection of coherently organized and interconnected concepts (Kinchin et al., 2000). However, obtaining quantitative measurements of such conceptual structures is not straightforward. Given that students' conceptual structure of a subject influences their learning and academic performance (Nesbit & Adesope, 2006), it is critical to find ways of measuring how the structure of domain knowledge develops with increasing expertise.

In the rest of the Introduction, we first briefly discuss the importance of measuring the knowledge structure of learners and consider how educational psychologists have attempted to capture these knowledge structures. We then review the ways in which cognitive psychologists measure and quantify semantic memory structure, with a focus on studies which use network science approaches. Finally, we focus on specific measures of network science used to quantify network structure and outline the study hypotheses.

✉ Cynthia S. Q. Siew
cynthia@nus.edu.sg

[1] Department of Psychology, National University of Singapore, 9 Arts Link, Block AS4, Singapore 117570, Singapore

Knowledge representation of experts and novices

A core principle in expertise research is that the underlying knowledge representations of experts and novices are different (Chi, 2006; Gobbo & Chi, 1986; Persky & Robinson, 2017; Wolff et al., 2017). Specifically, experts are said to understand the deep structure of problems, and hence are able to flexibly use this deep structure to solve a wide variety of problems within their domain of expertise. On the other hand, novices tend to focus on superficial aspects of the problem (Salkowski & Russ, 2018). For instance, classic studies in cognitive psychology demonstrated that expert chess players have better memory for meaningful chess formations than novices do (Chase & Simon, 1973), and experts uncover the deep structure of physics problems, whereas novices fixate on surface characteristics of the problem (Chi et al., 1981). In nonacademic domains, expert rock climbers' more accurate representation of action sequences and movements appear to underlie their better performance at cognitive tasks like remembering sequences of holds and moves (Whitaker et al., 2020).

The idea that experts have more sophisticated cognitive structures is highly relevant to the domain of education. An important goal of education is to develop learners into flexible and adaptive problem solvers. Ideally, students learn about more than just a collection of disparate facts about a given topic. Their own conceptual structure of a domain of knowledge should reflect the deep, hierarchical, and interconnected structure of that particular domain (Disessa & Sherin, 1998; Linn, 2006). Given that students' conceptual structure of an academic subject influences their subsequent learning and academic performance (Driver & Erickson, 1983; Nesbit & Adesope, 2006), it would be useful to have some way of measuring the conceptual representations of learners. However, how would one go about measuring such representations?

The most straightforward approach is to simply ask students to depict their domain knowledge through concept mapping (Novak, 2010). Students list discrete concepts and then connect concepts by indicating a line between related ideas. Typically, concept maps are qualitatively evaluated for their visual characteristics that are thought to reflect expert-like organization (e.g., a "star" shape) or novice-like organization (e.g., a linear, sequential structure; Kinchin et al., 2000). More recently, researchers have attempted to quantitatively evaluate these concept maps as mathematical graphs (Koponen & Nousiainen, 2014; Koponen & Pehkonen, 2010; Siew, 2018) so that methods from network science could be used to gain further insights into learners' conceptual representations. It is worth noting that the goal of investigating the *organization* of knowledge is somewhat different from summative assessments such as quizzes or tests that attempt to measure the existence or application of that knowledge (e.g., retrieval-based learning; Roediger & Karpicke, 2006). Specifically, summative assessments typically focus on quantifying test performance (e.g., how many questions is the student able to answer correctly), which is taken as a proxy for knowledge that a learner possesses, whereas our approach emphasizes the underlying *nature* of that knowledge: How are concepts associated with a particular knowledge domain interrelated and organized within the learner's long-term memory?

Taken together, the evidence suggests that the organization of knowledge among experts of a particular domain is a key factor for their better performance. This means that it is important to develop ways of measuring and quantifying knowledge structures so that we can delineate what is specifically different between experts and novices, and perhaps gain new ways of closing the gap between students who have a better or poorer grasp of the subject domain. In this paper we make use of a commonly used task in cognitive psychology (i.e., the semantic fluency or verbal fluency task) and methods from network science to uncover the conceptual structure of a learner's knowledge.

Measuring knowledge structure with network science

Siew (2020) suggested that methods from cognitive psychology used to study semantic memory retrieval and representation could be potentially adapted to study the properties of expert and novice conceptual structures within the education sciences. Semantic memory is the part of long-term memory that stores facts and information about the world and is commonly conceptualized as a network of concepts that are connected based on associative relationships or shared features (Collins & Loftus, 1975; Smith et al., 1974). Knowledge about a given domain could also be conceptualized in a similar manner as students learn about discrete facts or information that are organized in some meaningful way in their long-term memory. If we assume that such knowledge can be reasonably represented as a network structure of nodes and edges, where *nodes* represent distinct knowledge units or concepts and *edges* (or links) depict a relationship between pairs of concepts, it is possible to leverage on recent advancements within the field of cognitive network science to gain new insights into knowledge representations (for a review of cognitive network science, see Siew et al., 2019).

In the cognitive sciences, there is a lot of interest in measuring and quantifying semantic memory structure as it leads to greater insights into topics ranging from vocabulary acquisition (Beckage & Colunga, 2019; Hills et al., 2009) to cognitive decline (Wulff et al., 2019). To study the large-scale structure of semantic memory, cognitive scientists use network science approaches to analyze behavioral data from classic psychological tasks. For instance, there are citizen science projects that collect free associations (i.e., what are the first three words that come to mind for the concept "dog"?) through an online word association game—these free associations are commonly analyzed as semantic networks (De Deyne et al., 2019; Dubossarsky et al., 2017). Others have adopted a snowballing approach to collect free associations to estimate the semantic network structure of individuals (Morais et al., 2013; Wulff et al., 2021). Others have used the verbal fluency or semantic fluency task (i.e., name as many members of the "animal" category) to estimate semantic network structures (Borodkin et al., 2016; Kenett et al., 2013).

Semantic fluency networks

In the semantic fluency task, participants generate as many category members as they can within a short period of time. A characteristic pattern in fluency responses is that people tend to list related concepts in close proximity, leading to clusters of closely related responses (Troyer et al., 1998). For instance, for the category of "animals," a fluency list could be "dog, cat, mouse, pig, horse, cow." The first three responses are house pets, whereas the next three responses are farm animals. Various network estimation methods are used to infer the underlying network structure based on the observed fluency data by assuming that a simple random walk mechanism is producing the fluency responses (for a detailed review, see Zemla & Austerweil, 2018).

Network science metrics, such as average shortest path length (ASPL; which indexes the number of steps needed to traverse the network), clustering coefficient (CC; which indexes the extent of local clustering in the network), and modularity (Q; which indexes the presence of community structure or subclusters in the network), can be used to characterize semantic network structure. Communities are subclusters of nodes in the network that are more interconnected within themselves, and less interconnected with nodes outside the community. For example, Kenett et al. (2014) found that the semantic network structure of individuals with higher creative ability was less hierarchical and less spread out as compared with individuals with lower levels of creative ability. In another paper comparing the semantic networks of monolingual and bilingual speakers of Hebrew, Borodkin et al. (2016) found that the semantic network of bilinguals showed greater levels of local connectivity and less modularity, suggesting that L2 vocabulary structure was organized differently from L1 vocabulary.

In this paper, we use the semantic fluency task to estimate underlying the network structure of our participants for two reasons. First, the semantic fluency task is quick and easy to administer as compared with the free association task which requires the administering of hundreds if not thousands of cue words. An adapted version of the semantic fluency task could be reasonably administered in a classroom setting to take snapshots of students' learning. Second, much recent work has been devoted to the development and validation of different network estimation methods from fluency data (Christensen & Kenett, 2019; Zemla & Austerweil, 2018). This has led to the availability of accessible toolkits and tutorials for analyzing such data (Zemla et al., 2020).

Aim of study and hypotheses

The aim of the present study is to use the semantic fluency task to quantify the knowledge representation of various academic subjects among undergraduate and high school students. It is presumed that undergraduates would be more of an "expert" in various academic subjects than high school students as the former group would have had more years of education and recently completed their high school education. If the fluency task is able to pick up group-level structural differences between the knowledge networks of these two groups, it suggests that the task could be used to quantify knowledge representations of individuals with greater or lower levels of expertise in an academic domain. Because previous studies have successfully analyzed semantic fluency data as semantic networks (e.g., Borodkin et al., 2016; Kenett et al., 2016), we wanted to apply the same techniques to a different domain (i.e., academic subjects) and to a participant population not commonly studied in the psychological literature.

We expected to find structural differences between high school and undergraduate group-level semantic networks inferred from subject-specific fluency data. The undergraduate networks should have characteristics of a more sophisticated, well-integrated knowledge structure. This is where the quantification of network structure using commonly used network science measures such as average shortest path length (ASPL), clustering coefficient (CC), and modularity (Q) can be particularly useful. Specifically, we expected the undergraduate networks to have lower ASPLs than high school networks, suggesting greater overall navigability of the network as fewer steps/shorter paths are required to move between different concepts in their network. We also expected the undergraduate networks to have lower Qs than high school networks, suggesting higher level of interconnectivity across subclusters or subdomains of knowledge. As it was not immediately clear if expert networks would display higher or lower levels of local clustering among concepts, we did not have a specific prediction for CC. As for fluency networks of general semantic categories (i.e., animals and fruits), we did not expect to find differences as we reasoned that these should simply reflect domain-general knowledge that should be broadly similar across both groups.

Method

Participants

Fifty-five students (36 males, mean age = 16.7 years, *SD* = 1.03) were recruited from the National University of Singapore High School of Mathematics and Science (NUSH). The students' grade levels correspond to 10^{th}, 11^{th} and 12^{th} grades based on the North American high school system. Seventy-nine undergraduates (13 males, mean age = 21.1 years, *SD* = 1.59) were recruited from the National University of Singapore (NUS) and received either monetary reimbursement or credit towards fulfilling a course requirement for their participation. Thirty-five of the undergraduates were currently enrolled in the introductory psychology module, 15 had completed it during the course of their university education, one was a psychology minor, and remaining participants were psychology majors. All participants did not have any self-reported speech or language disorders. The study was approved by the Department of Psychology's Ethics Review Committee at the National University of Singapore.

Materials and procedure

All participants completed the general semantic fluency lists first, followed by subject-specific semantic fluency lists. For the general semantic fluency lists, participants were asked to generate as many category members for the general semantic categories of *animals* and *fruits*. The order of the cue words presented was randomized for each participant. Animals and fruits are well-established general semantic categories widely used across studies of semantic fluency (e.g., Lonie et al., 2009; Zemla et al., 2020). Following this, participants completed the subject-specific semantic fluency lists where they were asked to generate as many concepts (in the form of single words or short phrases) that were related to a specific academic subject (i.e., biology, chemistry, mathematics, physics, psychology). For example, students presented with the cue word "biology" responded with "cell," "life cycle," or "body." The order of presentation of subject-specific cue words was pseudo-randomized, the only constraint was that the cue word "psychology" could appear only after the first cue word.

Participants were given a maximum of 2 minutes to generate as many as responses as they could for each cue word. The procedure was as follows: the cue word was presented on a screen and participants entered their responses into a text box one at a time. Each response was programmed to fade from the screen after around 800 ms to avoid memory cuing from previous responses. Participants were instructed to refrain from repeating responses within each semantic fluency list or listing items with the same suffix (e.g., "cell" and "cells"). Due to the COVID-19 pandemic, data collection was conducted remotely via a custom-made web application.

Network estimation methods

Semantic fluency networks were constructed from the semantic fluency data and analyzed using a suite of R libraries specially developed for semantic network analysis (SemNA; Christensen & Kenett, 2019). SemNA contains three R packages (SemNetDictionaries, SemNetCleaner, and SemNet) and offers a single pipeline for preprocessing, estimating, and analyzing semantic networks from semantic fluency data. The analyses reported in this paper were conducted using the following versions of the R packages: SemNetDictionaries (Version 0.1.9), SemNetCleaner (Version 1.3.4), and SemNet (Version 1.4.4). First, fluency responses were cleaned using SemNetDictionaries and SemNetCleaner. The fluency data were first spell-checked and autocorrected using dictionaries, following which duplicate or inappropriate responses were removed. Finally, monikers or slightly different entries referring to the same concept were converted into a single, consistent label. For example, for the psychology lists, the responses "mental illness" and "mental disorder" were standardized as "mental disorder." This process was carried out via a combination of automated and manual effort.

Thereafter, cleaned regular response matrices and binary response matrices were produced for all responses belonging to each fluency list, from which semantic fluency networks were constructed. Each regular response matrix consisted of participant numbers running down the rows, with each participant's responses for that particular fluency list running across each row. Each binary response matrix consisted of participant numbers running down the rows, with columns representing unique responses across all participants for that particular fluency list. If participants had provided responses in particular columns, cells in the corresponding column were marked as "1"; if they did not, cells were marked as "0."

In semantic fluency networks, individual nodes represented concepts or members of a certain category in memory, and edges represented the existence of an association between pairs of concepts (Borodkin et al., 2016; Kenett et al., 2016). As discussed earlier, because fluency responses that occur close to each other also tend to be semantically related, it is possible to construct the semantic network given these inferred associations, and several methods have been developed for doing so (Christensen & Kenett, 2019; Zemla & Austerweil, 2018). The SemNet package contains four network estimation methods: Community Network (CN), Naïve Random Walk (NRW), Pathfinder (PF), and Correlation-based networks (CR). Please see Appendix: Table 8 for a complete list of abbreviations used to refer to the network estimation methods and network measures discussed in this paper. We performed our analyses using each of the available methods as each of them offer different advantages (for additional details, see Zemla & Austerweil, 2018). It was also important to ascertain if different methods would converge on similar network representations given the same data. Below, we briefly describe each network estimation method used to generate the semantic networks.

Naïve random walk networks The naïve random walk operates on the assumption that responses are generated by taking an uncensored random walk on a semantic network (Jun et al., 2015; Lerner et al., 2009). Under this assumption, adjacent responses are more likely to be semantically closely related. An edge is created between each pair of adjacent responses in the semantic fluency list. A co-occurrence matrix is created whereby the number of co-occurrences of adjacent pairs across the regular response matrix is counted and a threshold is applied to it such that only adjacent pairs that co-occur at least as many times as the threshold are estimated to have an edge in the network. In the present analysis, a threshold of 3 was used as it was the default value in SemNet; our results were similar across thresholds of up to 4.

Community networks In the community network method, a semantic network is estimated using a co-occurrence matrix where two responses that are not necessarily adjacent but occur within a fixed window or distance from each word is counted as a co-occurrence (Goñi et al., 2011). This co-occurrence matrix was derived from the cleaned regular response matrix for each fluency list. In the analyses reported below we used a window size of 2, as it was the default value in SemNet; our results were similar across different window sizes. These co-occurrences were then counted to infer if they were likely to have occurred by random chance. Using a 95% confidence interval from a binomial distribution, edges were created between responses that were significantly unlikely to occur together by random chance.

Pathfinder networks A binary response matrix was created whereby each row contained each participant's responses for a list and the columns contained unique fluency responses across the participants for that list. Each cell was denoted with "0" if the participant did not provide that response and "1" if they had done so. A proximity matrix between every pair of responses in the binary response matrix was then computed (Schvaneveldt, 1990). Only the path with the shortest distance between every pair of nodes was retained, and the network was then estimated by using a set of edges that links all nodes in the network while minimizing the total distance of all edges in the network.

Correlation-based networks Correlation-based network methods use a binary response matrix to estimate a semantic network based on co-occurrences of responses. An association measure between every pair of responses was computed using cosine similarity, producing an association matrix. As per common practice, only responses that were provided by at least two participants in each group were retained, in order to reduce spurious associations. Following that, responses were equated such that each group only contained the responses that were given by all other groups, which ensured that all groups had the same number of nodes, reducing confounding effects when making comparisons between networks (e.g., Borodkin et al., 2016). Note that this response equating step was only done for the correlation-based network estimation procedure in accordance with previous work that used this approach, and not for the other three network estimation methods. The Triangulated Maximally Filtered Graph (TMFG) method was implemented to connect the nodes in the network by maximizing the strength of their association to other nodes while making sure that the network is planar.

Network measures

After constructing networks from the various network estimation methods, we computed global network measures of the average shortest path length (ASPL), clustering coefficient (CC) and modularity index (Q) for each network to characterize its overall structure. The ASPL measures the average of the smallest number of steps needed to get from one node to another and is a measure of how efficiently a network can be navigated (Kleinberg, 2000). The CC refers to the probability that two neighbors of a randomly chosen node will also be neighbors of each other and is considered to be a measure of the level of local clustering in the network (Watts & Strogatz, 1998). The modularity of the network, Q, is a measure of the density of connections within communities as compared with the density of connections between communities (Fortunato, 2010; Newman, 2006). It also indicates the quality of the community partitions; networks with higher values of Q have robust or well-defined community structure where the density of connections within communities is much higher as compared with the density of connections between communities (Newman, 2006). The Louvain method (Blondel et al., 2008) was used to compute modularity.

Network analyses

Each network estimation method was used to estimate a semantic network for the NUS and NUSH groups for all seven fluency lists (i.e., *Animals, Fruits, Biology, Chemistry, Mathematics, Physics, Psychology*), resulting in 56 separate networks (7 lists × 2 groups × 4 methods). Two types of analyses, random network analyses and bootstrap analyses, were employed to statistically compare the networks generated from the fluency data. In the first approach, random networks with the same number of nodes and edges as the estimated networks were generated, in order to determine if the network measures observed are indeed significantly different from what would be expected from randomly generated networks with similar properties (e.g., Beckage et al., 2011). Random networks were generated such that the number of connections to each node, or the degree sequence, was preserved so that the general structure of the network was maintained for a fair comparison. For each simulated random network, global network measures of CC, ASPL, and Q were computed. This process was repeated 1,000 times for each network "type" (i.e., for all seven fluency lists, each student group [NUS and NUSH], and for all four network estimation methods). Global network measures (ASPL, CC, Q) were then computed for each group's random networks, resulting in random reference distributions. A one-sample *t* test was conducted to see if network measures of the estimated networks were significantly different from the distribution of network measures of simulated random networks. A significant result would indicate that the structure of the semantic networks cannot be easily "recovered" through a random generation process.

The second type of analysis, the bootstrap method (Efron, 1979), investigated if differences between the NUS and NUSH networks were statistically significant. We adopted a case-wise bootstrap approach whereby *N* number of participants were randomly sampled with replacement from the respective groups (where *N* was the number of participants). If two networks are indeed different from each other, then "partial" networks derived from a subset of participants should also be different from each other. For each sample, the same network estimation method was applied and global network measures (i.e., ASPL, CC, and Q) of that network were computed. This process was iterated 1,000 times for each network type, and global network measures for each group were then statistically compared using an analysis of covariance (ANCOVA) with number of edges included as a covariate to control for differences in network size. This was important to help control for confounds that might affect comparison between network measures of groups, as network measures can change depending on the ratio of edges to nodes of networks (van Wijk et al., 2010).

Results

Number of responses

Table 1 contains descriptive statistics for the number of semantic fluency responses and overall number of unique responses provided by each group of students for each cue word. For the chemistry and physics categories, there was no significant difference in the number of responses generated by NUS

Table 1 Descriptive statistics for fluency responses

	NUS				NUSH			
	M	*SD*	Range	Unique Response	*M*	*SD*	Range	Unique Response
Animals	31.98	7.25	14-44	292	25.41	12.5	3-50	395
Fruits	23.00	5.22	10-42	109	20.84	7.95	3-43	179
Psychology	19.05	6.33	6-38	581	1.35	6.34	3-27	305
Mathematics	21.28	8.35	7-62	440	18.25	6.98	3-36	329
Biology	21.54	8.87	6-43	580	18.41	10.0	2-50	403
Chemistry	20.03	8.09	4-42	465	17.94	7.59	2-36	385
Physics	18.86	11.7	3-86	401	19.00	8.18	3-45	353

NUS = National University of Singapore students; NUSH = NUS High School students.

and NUSH students, both $ps > .05$. For the biology and fruits categories, the results were marginally significant, $t(126) = 1.86$, $p = .07$ and $t(127) = 1.86, p = .07$ respectively, with NUS students producing slightly more words than NUS High students. For the animals, psychology, and mathematics categories, NUS students provided significantly more responses than NUS High students, $t(131) = 3.83$, $p = < .001$, $t(128) = 6.76, p < .001$, and $t(131) = 2.18, p = .03$, respectively.

For the general categories of fruit and animals, NUS students produced fewer unique number of responses overall, as compared with NUS High students. In all academic subject categories, NUS students generated more unique responses than NUS High students, with the largest difference found for psychology, where NUS and NUS High students provided 581 and 305 unique responses, respectively.

Global network measures of estimated networks

Global network measures of ASPL, CC and Q values were computed for the semantic networks of the NUS students and NUS High students for each of the seven fluency lists, with each of the four network estimation methods (community, naïve random walk, pathfinder, correlation-based networks). Table 2 shows the global network measures for the estimated semantic networks constructed using different network estimation methods. Generally, the estimated

Table 2 Global network measures for fluency networks constructed using different network estimation methods for each fluency list and each group

Network Measure	CN		NRW		PF		CR	
	NUS	NUSH	NUS	NUSH	NUS	NUSH	NUS	NUSH
				Animals				
ASPL	5.604	9.145	2.870	4.236	3.526	2.063	2.888	3.896
CC	0.366	0.151	0.231	0.084	0.653	0.744	0.758	0.707
Q	0.787	0.777	0.293	0.468	0.065	0.063	0.566	0.704
				Fruits				
ASPL	5.353	4.701	2.332	3.429	4.200	3.035	2.284	2.679
CC	0.135	0.121	0.469	0.184	0.658	0.699	0.776	0.747
Q	0.626	0.617	0.154	0.347	0.063	0.068	0.426	0.514
				Psychology				
ASPL	5.072	2.000	4.222	4.783	1.725	1.601	2.800	2.726
CC	0.097	0.000	0.062	0.051	0.812	0.829	0.717	0.725
Q	0.645	0.219	0.513	0.596	0.032	0.007	0.511	0.514
				Mathematics				
ASPL	7.224	2.156	4.180	3.887	2.212	1.921	2.769	3.438
CC	0.158	0.557	0.103	0.076	0.732	0.780	0.747	0.708
Q	0.713	0.375	0.434	0.470	0.048	0.019	0.583	0.635
				Biology				
ASPL	6.935	5.209	4.147	4.417	1.870	1.736	3.760	3.380
CC	0.211	0.046	0.076	0.056	0.774	0.792	0.710	0.720
Q	0.781	0.653	0.500	0.533	0.032	0.020	0.656	0.642
				Chemistry				
ASPL	6.549	7.047	3.679	4.194	1.934	1.669	3.479	3.286
CC	0.262	0.357	0.108	0.048	0.769	0.787	0.709	0.723
Q	0.743	0.666	0.461	0.537	0.028	0.030	0.643	0.675
				Physics				
ASPL	4.736	2.457	3.744	4.083	1.965	1.795	3.263	3.097
CC	0.213	0.223	0.096	0.064	0.774	0.770	0.724	0.729
Q	0.611	0.321	0.424	0.489	0.037	0.019	0.617	0.616

CN = community network; NRW = naïve random walk; PF = pathfinder; CR = correlation-based networks; NUS = National University of Singapore students; NUSH = NUS High School students; ASPL = average shortest path length; CC = clustering coefficient; Q = modularity.

NUS networks had a higher ASPL and higher Q than estimated NUS High networks, while there was no clear trend for CC. We note that this pattern runs counter to that of the bootstrapping analyses (see below) but wish to focus on the results of the bootstrapping analyses instead as the bootstrapped networks provide a further indication of the level of variability or uncertainty surrounding the network measures given multiple samples of the data. On the other hand, the analysis thus far only provides a single point estimate of the network measures as seen in Table 2.

Random network analyses

Global network measures of ASPL, CC and Q computed for each semantic network was compared against the same network measures obtained from the random network simulations. Recall that the goal was to assess if the estimated semantic networks are significantly different from that of a random network comprising of the same number of nodes and edges. The analysis revealed that most of the estimated semantic networks were indeed significantly different from the random network simulations, with very few exceptions (see Table 3; see also Appendix: Table 6 for results of the Bayesian analysis). This indicates that the structure of the estimated semantic networks is indeed meaningful (i.e., non-random) and cannot be recovered simply from chance.

Bootstrapping analyses

In order to test if differences between the semantic networks of the two groups are statistically significant, bootstrapping analyses were conducted to obtain bootstrapped network distribution parameters, with 1,000 iterations for each network type. Following that, bootstrapped network measures were statistically compared using an analysis of covariance (ANCOVA), with the number of edges included as a covariate to assess whether bootstrapped network measures significantly vary between the NUS and NUSH students. Across the bootstrapped networks, almost all comparisons were statistically significant (see Table 4; see also Appendix: Table 7 for results of the Bayesian analysis), with the following exceptions: CC for chemistry using correlation-based networks and ASPL for mathematics using the naïve random walk method.

In general, across most categories and network estimation methods, NUS bootstrapped networks had a significantly lower Q and lower ASPL as compared with NUSH bootstrapped networks. However, this pattern of lower Q and lower ASPL for the NUS networks did not hold when the networks were estimated using the pathfinder method. The pattern of results for CC varied depending on the network estimation method. While results from the community network and pathfinder methods indicated that NUS networks had a significantly lower CC than NUSH networks, the naïve random walk and correlation-based network methods indicated that NUS networks had a significantly higher CC than NUSH networks.

The pattern of results for general categories (i.e., animals and fruits) are similar to that of the subject-specific categories (i.e., biology, chemistry, mathematics, physics, and psychology) across different network estimation methods, except for pathfinder networks. For pathfinder networks, while NUS networks had a larger ASPL and lower CC than the NUSH networks across most categories, NUS networks had a lower Q for general categories but a higher Q for subject-specific categories (excluding psychology).

To provide a qualitative comparison of network structures, network visualizations of the estimated fluency networks are provided in Fig. 1. Figure 1 depicts a variety of networks from each cue word, estimated using different network estimation techniques, and includes network measures for each network for easy comparison. Overall, the semantic networks of the NUS students appear to be visually less spread out and compartmentalized than that of the NUS High students, reflecting the lower Q and lower APSL of the NUS networks. This pattern is broadly consistent across various estimation methods and cue words.

General discussion

The goal of the present study was to analyze fluency responses as semantic networks of concepts to quantify the knowledge structure of students with relatively higher and lower levels of expertise in those subjects (i.e., undergraduate students who have relatively more "expertise" than high school students). Participants were asked to generate concepts related to different academic subjects (e.g., psychology and biology) and their fluency responses were analyzed as semantic networks whose structure were further quantified using network science metrics.

The results were somewhat consistent with our hypotheses. We expected to find no difference in the domain-general networks (i.e., animals, fruits) of NUS and NUS High students. We also expected that the structure of domain-specific networks (i.e., biology, physics, chemistry, mathematics, psychology) to reflect more efficient, well-integrated organization for the group with more expertise on those subjects (i.e., NUS undergraduates). In our analysis, we found that across both general and domain-specific cue words, bootstrapped NUS networks had lower ASPL and lower Q than bootstrapped NUS High networks. This pattern of finding was largely consistent across networks estimated using different network estimation methods (except for pathfinder networks). For the CC measure, there was not a consistent pattern across cue words and network estimation methods.

Table 3 Summary of results from the random network analysis

Network Measures	CN		NRW		PF		CR	
	NUS	NUSH	NUS	NUSH	NUS	NUSH	NUS	NUSH
				Animals				
ASPL.M	3.985***	4.862***	2.787***	3.601***	1.987***	1.767***	2.780***	2.860***
ASPL.SD	0.085	0.326	0.019	0.035	0.003	0.001	0.038	0.026
CC.M	0.033***	0.028***	0.136***	0.044***	0.412***	0.519***	0.088***	0.064***
CC.SD	0.011	0.021	0.005	0.004	0.002	0.001	0.011	0.009
Q.M	0.548***	0.658***	0.265***	0.436***	0.067***	0.052***	0.380***	0.387***
Q.SD	0.013	0.021	0.005	0.006	0.002	0.001	0.012	0.011
				Fruits				
ASPL.M	3.946***	3.952***	2.393***	2.957***	2.057***	1.921***	2.473***	2.487***
ASPL.SD	0.481	0.689	0.025	0.041	0.010	0.005	0.054	0.046
CC.M	0.053***	0.049***	0.309***	0.130***	0.381***	0.510***	0.150***	0.136***
CC.SD	0.047	0.065	0.011	0.010	0.004	0.004	0.020	0.019
Q.M	0.602***	0.583***	0.216***	0.342***	0.110***	0.087***	0.357***	0.361***
Q.SD	0.035	0.047	0.008	0.008	0.003	0.002	0.020	0.020
				Psychology				
ASPL.M	3.796***	1.658***	3.980***	4.462***	1.698***	1.692***	2.447***	2.447***
ASPL.SD	0.334	0.373	0.044	0.104	0.001	0.001	0.047	0.049
CC.M	0.057***	0.083***	0.033***	0.028***	0.556***	0.471***	0.133***	0.135***
CC.SD	0.038	0.276	0.004	0.006	0.001	0.001	0.022	0.021
Q.M	0.619***	0.356***	0.497***	0.595**	0.039***	0.041***	0.365***	0.364***
Q.SD	0.030	0.144	0.005	0.007	0.001	0.001	0.021	0.020
				Mathematics				
ASPL.M	4.354***	2.290***	3.510***	3.682***	1.773***	1.747***	2.639***	2.671***
ASPL.SD	0.341	0.308	0.039	0.053	0.001	0.001	0.040	0.033
CC.M	0.039***	0.178***	0.062***	0.052***	0.521***	0.492***	0.108***	0.088***
CC.SD	0.027	0.132	0.005	0.006	0.001	0.001	0.015	0.013
Q.M	0.634***	0.380	0.424***	0.476***	0.047***	0.047***	0.373***	0.380***
Q.SD	0.024	0.076	0.006	0.007	0.001	0.001	0.015	0.015
				Biology				
ASPL.M	4.709***	4.453***	3.856***	4.070***	1.736***	1.726***	2.781***	2.765***
ASPL.SD	0.256	0.696	0.036	0.057	0.001	0.001	0.030	0.032
CC.M	0.028***	0.038***	0.034***	0.031***	0.548***	0.491***	0.074***	0.079***
CC.SD	0.017	0.045	0.004	0.005	0.001	0.001	0.011	0.011
Q.M	0.660***	0.640***	0.473***	0.526***	0.039***	0.041***	0.386***	0.385***
Q.SD	0.018	0.036	0.005	0.006	0.001	0.001	0.013	0.013
				Chemistry				
ASPL.M	4.319***	3.609***	3.585***	4.024***	1.731***	1.707***	2.725***	2.727***
ASPL.SD	0.214	0.211	0.039	0.060	0.001	0.001	0.032	0.033
CC.M	0.034***	0.051***	0.055***	0.034***	0.537***	0.503***	0.080***	0.084***
CC.SD	0.019	0.036	0.005	0.005	0.001	0.001	0.011	0.012
Q.M	0.602***	0.527***	0.434***	0.524***	0.043***	0.044***	0.382***	0.381***
Q.SD	0.020	0.032	0.005	0.006	0.001	0.001	0.013	0.013
				Physics				
ASPL.M	3.630***	2.636***	3.509***	3.783***	1.735***	1.725***	2.712***	2.689***
ASPL.SD	0.413	0.324	0.042	0.053	0.001	0.001	0.038	0.035
CC.M	0.066***	0.143***	0.063***	0.044***	0.523***	0.516***	0.093***	0.093***
CC.SD	0.051	0.098	0.006	0.005	0.001	0.001	0.013	0.013
Q.M	0.572***	0.446***	0.427***	0.483***	0.046***	0.046***	0.382***	0.378***
Q.SD	0.038	0.058	0.006	0.006	0.001	0.001	0.015	0.014

CN = community network; NRW = naïve random walk; PF = pathfinder; CR = correlation-based networks; NUS = National University of Singapore students; NUSH = NUS High School students; ASPL = average shortest path length; CC = clustering coefficient; Q = modularity. M = mean, SD = standard deviation.

$^{**}p < .01$, $^{***}p < .001$.

What have we learned about knowledge representation?

Our analyses revealed that NUS networks had lower ASPL and lower Q than NUS High networks. Recall that ASPL is a global network metric that measures the average number of steps required to connect all pairs of concepts in the network. A lower ASPL value means that fewer steps are needed on average to connect from one concept to another. Hence, ASPL is an indicator of the ease of navigating across the entire network (Kleinberg, 2000). In the context of knowledge representation, the lower ASPL of the NUS networks may suggest that the knowledge representation of NUS students has a more easily or efficiently navigated structure than that of the NUS High networks.

In contrast to ASPL which is a macro-level network science measure, modularity, Q, is a meso-level measure that quantifies the quality of clusters or communities within the network (Newman, 2006). Within the context of knowledge representation, a lower value of Q indicates the presence of less segregated or less clearly defined communities of subtopics within a domain of knowledge. The lower Q of NUS networks as compared with NUS High networks suggests that NUS networks are less modular than the NUS High networks, such that there are more interconnections across sub-topics within the same domain of knowledge. Furthermore, the random network analyses indicate that the structure of the NUS and NUS High networks is "nonrandom." In other words, the observed network structure cannot be recovered simply through random re-associations of the same concepts. Taken together, it appears that the network analysis was able to characterize the emergence of "expert" structure: the bootstrapped NUS networks appeared to be both more integrated (lower Q) but also more efficiently structured (lower ASPL). These results demonstrate how semantic network analysis can provide researchers with a quantitative way of describing various aspects of the structure of semantic memory and domain knowledge that would not have been possible otherwise.

The impact of additional years of education

The finding that NUS networks have lower Q and ASPL than NUS High networks seems to resemble that of previous studies by Kenett and colleagues who reported similar differences between the semantic fluency networks of individuals with higher levels of creative ability and the semantic fluency networks of individuals with lower levels of creative ability. Specifically, they found that the semantic network of individuals with higher creative ability had lower Q and lower ASPL than the network of individuals with lower creative ability (Kenett et al., 2014; Kenett et al., 2016). In Borodkin et al.'s (2016) study, they reported that semantic fluency networks of L2 speakers had lower Q than the semantic fluency networks of L1 speakers. Given that we observed similar patterns for the undergraduate sample as compared with the high school sample, this may suggest that additional years of education, particularly within a university setting, could be important for developing flexible, abstract thought and promoting creative ability (Walker & Finney, 1999). Broadly, these observations are generally aligned with empirical evidence supporting the idea that creative ability levels are in fact closely linked to the individual's domain-specific knowledge and expertise levels, rather than their domain-general skills or traits (Baer, 1998, 2015). Perhaps a semantic network that is more "experienced" has the prerequisite structural features in which creative skill and ability could then emerge and develop from.

Diverse experiences associated with the college experience could also be another factor contributing to the flatter, less modular network structure of the NUS students. This could be a possibility given that previous work by Christensen et al. (2018) found that individuals with higher levels of openness to experience, a personality variable related to greater enjoyment and seeking out novel experiences, produced semantic networks that were more interconnected and flexible than individuals with lower levels of openness to experience. Another relevant example is given in work by Lydon-Staley et al. (2021), who showed that the curiosity disposition of participants was associated with different information seeking styles that led to structural differences in their knowledge networks. Specifically, participants who were more of a "busybody" explored diverse ideas, resulting in networks with longer path lengths and lower local clustering, whereas participants who were more of a "hunter" explored in a way that filled in gaps in their knowledge, resulting in networks with shorter path lengths and higher clustering (Lydon-Staley et al., 2021). One might expect university students in our study to have more of a "hunter" disposition, where they are beginning to specialize in domains that are of interest to them, and to some extent this is indeed reflected in the lower ASPL of their fluency networks as compared with the networks of high school students. Finally, it is important to emphasize that we are not claiming a causal relation between education level and network structure. We simply wished to point out parallels between the current finding and previous findings using a highly similar methodology. Longitudinal studies of individual semantic fluency networks across the learner's education would be necessary for establishing such causal relations. Future work could potentially take inspiration from the extensive literature on the development and aging of semantic networks (Dubossarsky et al., 2017; Steyvers & Tenenbaum, 2005) to obtain deeper insights into the development of domain-specific semantic networks.

Table 4 Summary of results from the bootstrapping network analysis

	CN				NRW				PF				CR			
	NUS *M* (*SD*)	NUSH *M* (*SD*)	*p*	partial eta sq	NUS *M* (*SD*)	NUSH *M* (*SD*)	*p*	partial eta sq	NUS *M* (*SD*)	NUSH *M* (*SD*)	*p*	partial eta sq	NUS *M* (*SD*)	NUSH *M* (*SD*)	*p*	partial eta sq
Animals																
ASPL	3.355 (0.117)	4.293 (0.328)	<.001	0.736	2.985 (0.064)	4.464 (0.226)	<.001	0.583	2.873 (0.143)	1.895 (0.076)	<.001	0.286	3.286 (0.291)	4.603 (0.471)	<.001	0.737
CC	0.307 (0.019)	0.387 (0.02)	<.001	0.793	0.193 (0.015)	0.067 (0.01)	<.001	0.638	0.71 (0.011)	0.768 (0.012)	<.001	0.047	0.73 (0.009)	0.691 (0.007)	<.001	0.83
Q	0.537 (0.021)	0.631 (0.029)	<.001	0.724	0.328 (0.012)	0.521 (0.018)	<.001	0.724	0.038 (0.007)	0.043 (0.009)	<.001	0.035	0.632 (0.028)	0.709 (0.018)	<.001	0.768
Fruits																
ASPL	3.524 (0.264)	4.144 (0.388)	<.001	0.552	2.334 (0.06)	3.498 (0.163)	<.001	0.788	3.877 (0.358)	2.563 (0.178)	<.001	0.208	2.419 (0.11)	2.801 (0.198)	<.001	0.705
CC	0.245 (0.037)	0.359 (0.029)	<.001	0.587	0.412 (0.034)	0.163 (0.016)	<.001	0.763	0.696 (0.014)	0.709 (0.015)	0.175	0.00092	0.755 (0.008)	0.724 (0.011)	<.001	0.759
Q	0.559 (0.032)	0.628 (0.034)	<.001	0.565	0.176 (0.01)	0.371 (0.02)	<.001	0.856	0.043 (0.009)	0.061 (0.013)	<.001	0.267	0.47 (0.028)	0.542 (0.029)	<.001	0.702
Psychology																
ASPL	4.344 (0.266)	5.299 (0.697)	<.001	0.015	4.6 (0.19)	4.942 (0.299)	0.037	0.002	1.639 (0.04)	1.645 (0.028)	<.001	0.499	2.744 (0.257)	2.963 (0.363)	<.001	0.006
CC	0.425 (0.017)	0.463 (0.029)	<.001	0.073	0.049 (0.007)	0.034 (0.009)	0.138	0.001	0.813 (0.008)	0.851 (0.006)	<.001	0.484	0.7 (0.026)	0.684 (0.029)	<.001	0.049
Q	0.644 (0.023)	0.705 (0.029)	<.001	0.062	0.564 (0.014)	0.628 (0.018)	<.001	0.009	0.028 (0.005)	0.008 (0.002)	<.001	0.403	0.522 (0.043)	0.562 (0.04)	<.001	0.051
Mathematics																
ASPL	4.238 (0.343)	4.343 (0.314)	0.001	0.005	4.368 (0.385)	4.13 (0.132)	0.064	0.002	1.997 (0.078)	1.8 (0.049)	<.001	0.723	3.193 (0.243)	3.512 (0.252)	<.001	0.058
CC	0.392 (0.019)	0.421 (0.025)	<.001	0.152	0.089 (0.01)	0.064 (0.011)	<.001	0.055	0.776 (0.01)	0.808 (0.007)	<.001	0.747	0.716 (0.011)	0.698 (0.01)	<.001	0.241
Q	0.626 (0.024)	0.657 (0.027)	<.001	0.018	0.471 (0.015)	0.516 (0.016)	<.001	0.054	0.033 (0.007)	0.02 (0.004)	<.001	0.515	0.605 (0.025)	0.64 (0.021)	<.001	0.084
Biology																
ASPL	4.244 (0.215)	4.624 (0.359)	<.001	0.044	4.451 (0.151)	4.734 (0.2)	<.001	0.017	1.734 (0.044)	1.681 (0.036)	<.001	0.544	3.76 (0.373)	4.133 (0.379)	<.001	0.013
CC	0.4 (0.016)	0.429 (0.024)	<.001	0.236	0.063 (0.007)	0.043 (0.008)	<.001	0.042	0.795 (0.009)	0.823 (0.007)	<.001	0.538	0.698 (0.012)	0.683 (0.014)	<.001	0.188
Q	0.631 (0.021)	0.671 (0.026)	<.001	0.077	0.547 (0.014)	0.591 (0.018)	<.001	0.058	0.029 (0.005)	0.019 (0.005)	<.001	0.156	0.662 (0.025)	0.68 (0.023)	<.001	0.041

Table 4 (continued)

	CN				NRW				PF				CR			
	NUS *M* (*SD*)	NUSH *M* (*SD*)	*p*	partial eta sq	NUS *M* (*SD*)	NUSH *M* (*SD*)	*p*	partial eta sq	NUS *M* (*SD*)	NUSH *M* (*SD*)	*p*	partial eta sq	NUS *M* (*SD*)	NUSH *M* (*SD*)	*p*	partial eta sq
Chemistry																
ASPL	3.959 (0.162)	4.527 (0.395)	<.001	0.162	3.888 (0.085)	4.582 (0.21)	<.001	0.201	1.825 (0.049)	1.671 (0.038)	<.001	0.826	3.708 (0.384)	3.837 (0.404)	<.001	0.012
CC	0.381 (0.018)	0.424 (0.021)	<.001	0.392	0.083 (0.01)	0.04 (0.008)	<.001	0.276	0.785 (0.008)	0.814 (0.008)	<.001	0.771	0.696 (0.012)	0.697 (0.011)	<.001	0.008
Q	0.608 (0.02)	0.663 (0.027)	<.001	0.206	0.497 (0.013)	0.574 (0.015)	<.001	0.231	0.025 (0.005)	0.022 (0.006)	<.001	0.098	0.649 (0.028)	0.665 (0.026)	0.301	0.000536
Physics																
ASPL	4.098 (0.217)	4.256 (0.255)	<.001	0.036	3.893 (0.108)	4.319 (0.148)	<.001	0.29	1.852 (0.06)	1.724 (0.038)	<.001	0.693	3.394 (0.299)	3.739 (0.362)	<.001	0.115
CC	0.38 (0.023)	0.406 (0.02)	<.001	0.287	0.084 (0.011)	0.051 (0.009)	<.001	0.295	0.787 (0.008)	0.803 (0.008)	<.001	0.493	0.706 (0.012)	0.693 (0.01)	<.001	0.204
Q	0.624 (0.024)	0.64 (0.024)	<.001	0.027	0.467 (0.015)	0.534 (0.015)	<.001	0.381	0.028 (0.005)	0.021 (0.004)	<.001	0.287	0.631 (0.025)	0.653 (0.023)	<.001	0.069

CN = community network; NRW = naïve random walk; PF = pathfinder; CR = correlation-based networks; NUS = National University of Singapore students; NUSH = NUS High School students; ASPL = average shortest path length; CC = clustering coefficient; Q = modularity; *M* = mean; *SD* = standard deviation; *p* = *p* value.

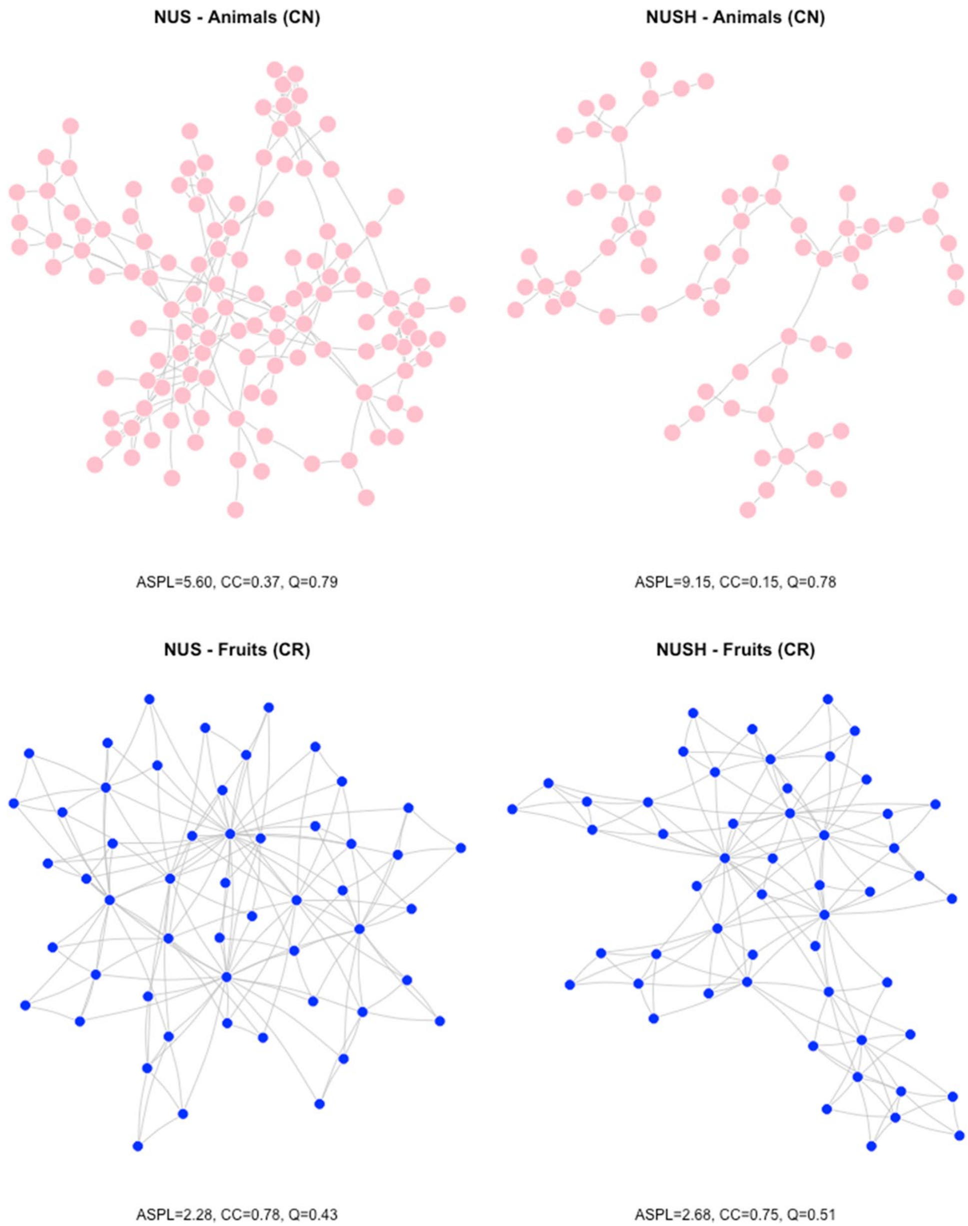

Fig. 1 Visualizations of semantic fluency networks for the NUS (left) and NUS High (right) groups. CN = community network; NRW = naïve random walk; PF = pathfinder; CR = correlation-based networks; NUS = National University of Singapore students; NUSH = NUS High School students; ASPL = average shortest path length; CC = clustering coefficient; Q = modularity

Considering the limits of the semantic fluency task

The original goal of presenting cue words that are more specific than general semantic categories was to examine the possibility that the semantic fluency task could be used to measure the structure of specific domain knowledge within educational settings. However, a conservative interpretation of the present findings is that the behavioral task and/or the network analysis was in fact *not* sensitive to the type of cue word (general or specific) used in the fluency task, since the same pattern of finding was observed across both general and domain-specific cue words—NUS networks tended to have lower ASPL and lower Q than NUS High networks.

Consider the very nature of the semantic fluency task, which is to quickly generate as many instances of concepts related to a semantic category or a domain of knowledge. Such a paradigm necessarily prioritizes quick retrieval of accessible concepts involving a fast and somewhat

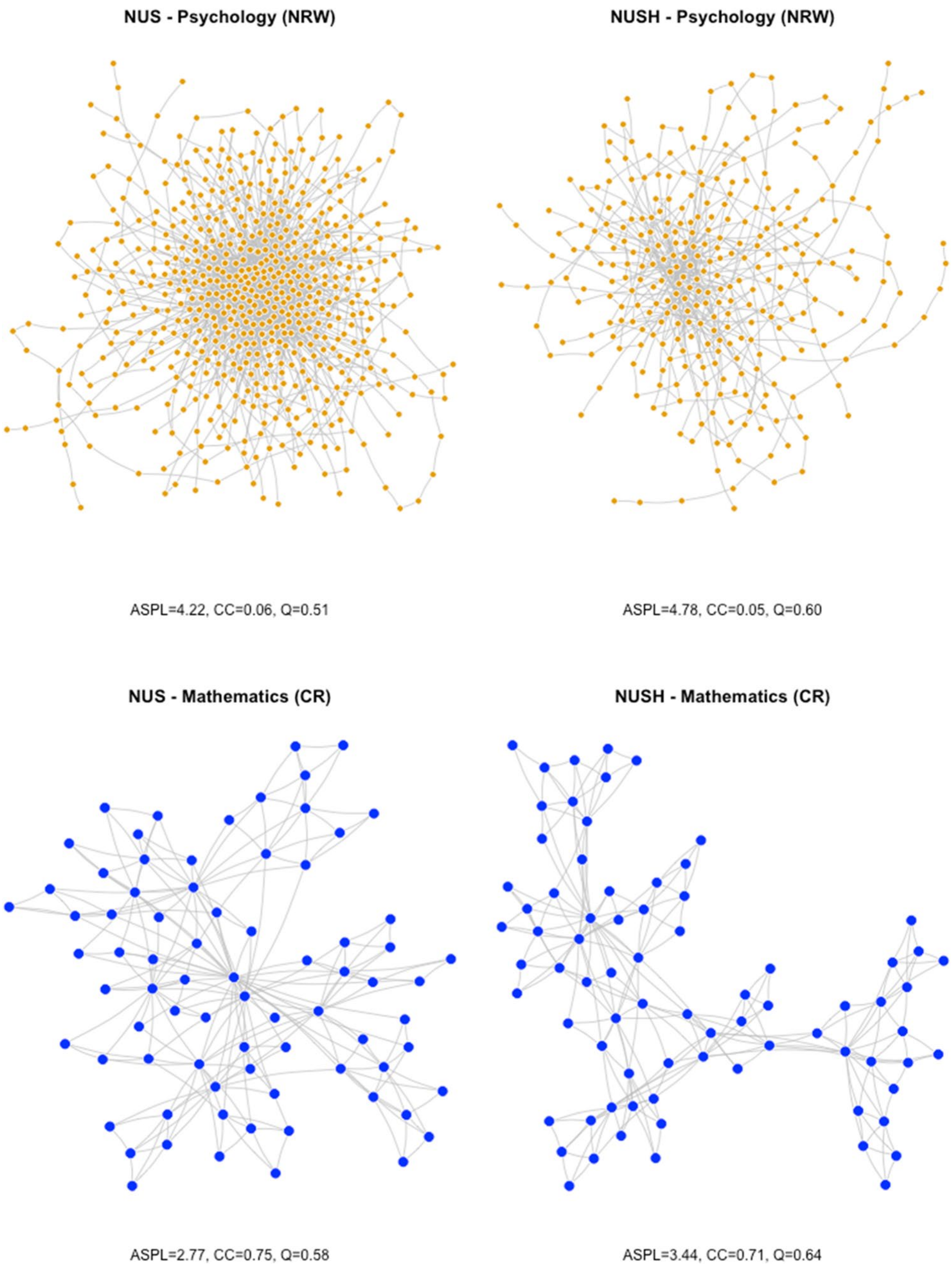

Fig. 1 (continued)

superficial search of memory, rather than executing a deep search of what one knows about a given topic. Hence, a possible interpretation is that the network structure inferred from fluency responses may reflect structural characteristics of semantic memory in general, rather than structural characteristics of specific knowledge domains. Some support for this claim is supported by a frequency analysis of participants' responses. As shown in Table 5, the responses that occurred most frequently tended to be general concepts related to the subject and were quite similar across both groups of participants. In other words, the similarity of the fluency response profiles for both groups could indicate that the semantic fluency task is primarily measuring general memory structure rather than specific knowledge representations. This notion is also supported by an analysis comparing the relative frequency of fluency responses from the two groups, where it is not obvious if either group is producing concepts that are more general or specific than the other group (see Appendix: Fig. 2). Additional work is necessary to further evaluate the advantages and challenges of using the semantic fluency task to measure domain-specific knowledge.

Fig. 1 (continued)

Comparison of various network estimation measures

An important methodological contribution of the present study lies in the application of four different network estimation techniques to derive semantic network structure from fluency data. Previous papers in this area typically use a single method for network estimation. Hence, our present results provide a starting point to compare general differences in the structure of the estimated networks from these techniques, and potentially offer some suggestions and recommendations for future work in this area.

An overview of the network measures in Table 2 (and Fig. 1) provides the following initial observations: (1) CN and CR networks tend to be smaller in size than PF and NRW networks, and (2) PF and CR networks tend to be relatively more "well-connected" than CN and NRW networks. These are in line with the two recommendations discussed by Zemla and Austerweil (2018). First, the estimated PF and NRW networks are much larger because they contain a node for each response in the data and both methods lead to the estimation of a single, fully connected network. On the other hand, the CN and CR methods do not have such constraints. Zemla and Austerweil (2018) point out that while

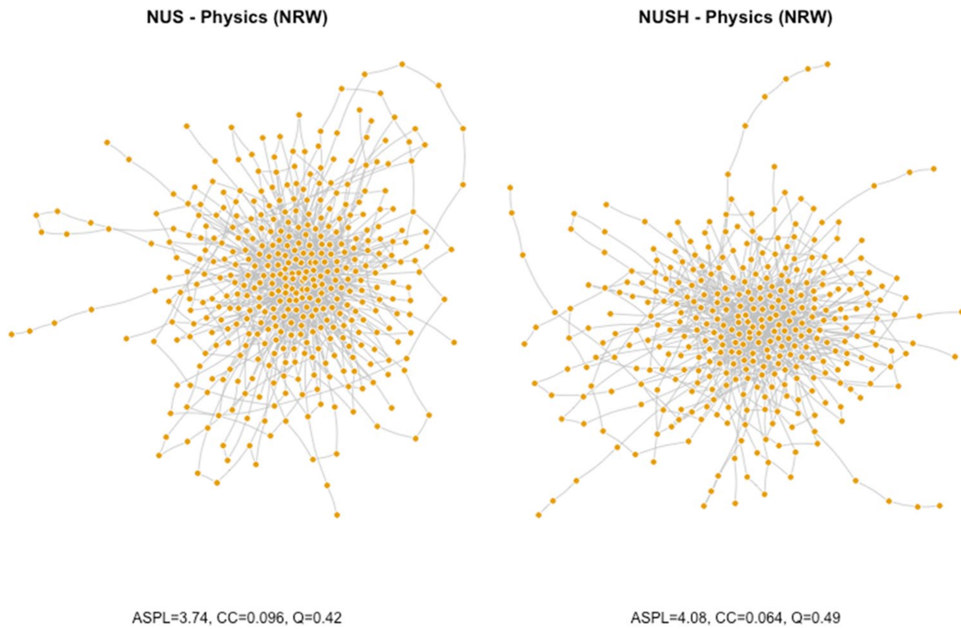

Fig. 1 (continued)

Table 5 Top 10 most frequently listed responses across psychology, biology, and animal networks

Psychology		Biology		Animals	
NUS	NUSH	NUS	NUSH	NUS	NUSH
brain	brain	cell	cell	cat	dog
behaviour	mental disorder	brain	plant	dog	cat
mind	depression	DNA	photosynthesis	lion	lion
Sigmund Freud	mind	heart	animal	tiger	tiger
abnormal psychology	mental health	plant	DNA	bear	fish
biology	thought	animal	evolution	pig	bird
cognitive	emotion	photosynthesis	ecology	rabbit	bear
development psychology	anxiety	mitochondria	genetics	cow	human
memory	behaviour	nucleus	respiration	giraffe	shark
statistics	neuron	organ	mitichindria	elephant	pig

NUS = National University of Singapore students; NUSH = NUS High School students.

this constraint may be important for certain assumptions in psychological modeling, it may also lead to the creation of several spurious edges. Second, the relatively more "well-connected" PF networks could be the result of a liberal edge creation criteria that minimized non-edge similarity, whereas the CN and NRW methods are relatively sparser due to a more conservative edge creation criteria that maximized edge similarity (see Zemla & Austerweil, 2018, for more details). Finally, CR networks tend to have relatively high local clustering, likely due to the way that the correlations of fluency response patterns are computed such that triplets of similar concepts tend to become easily connected in the resulting network.

In the present paper, we found consistent patterns in network differences between the NUS and NUSH groups for the CN, NRW, and CR methods, with the PF method being more of the odd one out. We suggest that PF networks may be too "hyper-connected" such that the presence of too many spurious edges (due to the constraint of having to include all responses and connect them all in a single network, plus a highly liberal edge creation approach) may in fact mask the underlying knowledge structure of the two groups. However, we also highlight that more work is needed to fully understand the strengths and weaknesses of each network estimation approach so that the researcher can properly evaluate the results of the network analysis in light of these strengths and weaknesses.

Limitations and future directions

In this final section we review and discuss the limitations of the current study. First, sample sizes were less than expected due to the COVID-19 pandemic limiting our data-collection efforts. In particular, constraints on data collection led to fewer numbers of high school students in our sample than expected. The somewhat large difference in sample sizes for high school students and university students could have implications on the results of the network estimations; however, we have conducted additional bootstrapping analyses where sample sizes of the two groups were matched and found that differences in the structure of NUS and NUSH networks are not merely a by-product of the larger sample sizes of the NUS group (for more details please see Appendix: Table 9). Nevertheless, it is important to emphasize that network estimation techniques are still in development and more research is needed to fully understand the impacts of unequal sample sizes. Given the current state of the art in semantic network analysis, we wish to simply highlight this potential limitation and point out that our raw data is available on OSF for re-analysis when updated techniques that can take unequal sample sizes into account become available.

Remote data collection also became necessary which tends to reduce data quality and prevented us from conducting additional tests of cognitive and memory processes. In an ideal lab-based setting, it would have been feasible to include an entire battery of cognitive and working memory tasks. This is particularly relevant as it is known that semantic fluency task performance is dependent on control processes and working memory span (Amunts et al., 2020; Shao et al., 2014; Whiteside et al., 2016). Nevertheless, we felt that the current investigation still provided a reasonable first foray into examining the potential of semantic network analysis on measuring knowledge representations of students of different levels of experience. This approach could complement qualitative approaches used to measure students' knowledge structures, such as concept mapping.

Given our findings of a clear effect of additional years of education on semantic network structure, future comparisons of groups of students using the subject-specific fluency task will need to exert a high level of control on participant characteristics to enable meaningful expert–novice comparisons. One possible future direction would be to compare student populations with the same level of education but who differ on their specific subject experience (e.g., same year undergraduates who have different degree specializations). Another direction is to conduct a longitudinal study where subject-specific fluency lists are completed at various stages of the student's educational or university career. Finally, it would also be important to investigate if the group-level differences we have observed in the present paper can also be replicated at the level of individual semantic networks. We wish to highlight that individual network analyses are possible with Zemla and Austerweil's (2018) *snafu* Python library, especially if each participant has provided multiple fluency lists per cue word and reasonable measures are taken to reduce the computational cost of conducting such an analysis.

Conclusion

In closing, we wish to briefly highlight how the present study relates to, and contributes toward, our understanding of human cognitive universals. The semantic fluency task provides one measure of how people navigate and search their semantic memory (Hills et al., 2012). A key observation from the literature is that the way humans search for information in cognitive spaces is strikingly similar to the way animals forage for resources in physical spaces (Hills et al., 2008). This suggests the presence of search or foraging mechanisms that are evolutionarily ancient. The present study builds on this idea by showing that generalized cognitive search mechanisms are likely involved when (both WEIRD and non-WEIRD) humans search in both domain-general (i.e., the semantic category of *animals*) as well as domain-specific knowledge spaces (i.e., the domain of *physics*).

To recapitulate, the current study found that the organization of general semantic memory and domain-specific knowledge representations was better connected and integrated for students with more years of education, and hence more of an "expert" in various academic domains. Although further work is needed to understand the limits of using the semantic fluency task to probe the deep structure of expert representations, the current study also established that it is in principle possible to adapt the semantic fluency task to quantify knowledge representations of more specific domains. Our results have implications for researchers interested in characterizing the specific nature of expert cognitive representations.

Appendix

Table 6 Bayes factors for random network analysis

Cue	Network Measure	CN.NUS	CN.NUSH	NRW.NUS	NRW.NUSH	PF.NUS	PF.NUSH	CR.NUS	CR.NUSH
Animals	ASPL	∞	∞	∞	∞	∞	∞	∞	∞
	CC	∞	∞	∞	∞	∞	∞	∞	∞
	Q	∞	∞	∞	∞	482.942	∞	∞	∞
Fruits	ASPL	∞	385.458	∞	∞	∞	∞	∞	∞
	CC	691.173	400.607	∞	∞	∞	∞	∞	∞
	Q	190.065	213.507	∞	157.612	∞	∞	∞	∞
Psychology	ASPL	∞	301.017	∞	∞	∞	∞	∞	∞
	CC	366.662	37.247	∞	∞	∞	∞	∞	∞
	Q	278.496	316.783	∞	-0.390	∞	∞	∞	∞
Mathematics	ASPL	∞	82.422	∞	∞	∞	∞	∞	∞
	CC	∞	∞	∞	∞	∞	∞	∞	∞
	Q	∞	-1.498	683.344	270.241	424.144	∞	∞	∞
Biology	ASPL	∞	385.450	∞	∞	∞	∞	∞	∞
	CC	∞	11.690	∞	∞	∞	∞	∞	∞
	Q	∞	61.977	∞	418.016	∞	∞	∞	∞
Chemistry	ASPL	∞	∞	∞	∞	∞	∞	∞	∞
	CC	∞	∞	∞	∞	∞	∞	∞	∞
	Q	∞	∞	∞	∞	∞	∞	∞	∞
Physics	ASPL	∞	129.436	∞	∞	∞	∞	∞	∞
	CC	∞	249.233	∞	∞	∞	∞	∞	∞
	Q	355.613	∞	136.280	344.452	∞	∞	∞	∞

We conducted Bayesian one-sample *t* tests to compare the network measures of 1,000 randomly generated networks against the network measures of the corresponding estimated networks. In all comparisons, we found that almost all Bayes factors were well above 100[1] (several were of infinite values), indicating that the network structure of the estimated networks indeed differed from the distribution of random networks generated with the same number of nodes and edges, in line with the results of the frequentist approach reported in the main text. In order to improve the presentation of Table 6, we report $\log_{10}$(Bayes factors).

[1]Bayes factor is defined as the ratio of the likelihood of the alternative hypothesis (i.e., the estimated network measure is different from the random network distribution) to the likelihood of the null hypothesis (i.e., the estimated network measure is not different from the random network distribution). Based on the recommendations by Lee and Wagenmakers (2014), a Bayes factor of 100 indicates very strong evidence for the alternative hypothesis.

Table 7 Bayes factors for bootstrapping network analysis

Cue	Network Measure	CN (LogBF10)	NRW (LogBF10)	PF (LogBF10)	CR (LogBF10)
Animals	ASPL	1585.228	3057.829	3023.450	1873.075
		NUS < NUSH	NUS < NUSH	NUS > NUSH	NUS < NUSH
	CC	1663.042	3301.982	2080.700	1984.934
		NUS < NUSH	NUS > NUSH	NUS < NUSH	NUS > NUSH
	Q	1589.979	3803.511	115.905	2140.695
		NUS < NUSH	NUS < NUSH	NUS < NUSH	NUS < NUSH
Fruits	ASPL	836.659	3145.994	1867.283	1580.171
		NUS < NUSH	NUS < NUSH	NUS > NUSH	NUS < NUSH
	CC	1392.183	3137.106	184.239	1565.482
		NUS < NUSH	NUS > NUSH	NUS < NUSH	NUS > NUSH
	Q	900.605	3685.277	630.557	1474.506
		NUS < NUSH	NUS < NUSH	NUS < NUSH	NUS < NUSH
Psychology	ASPL	614.579	380.465	816.959	782.808
		NUS < NUSH	NUS < NUSH	NUS < NUSH	NUS < NUSH
	CC	505.174	603.421	2075.389	85.708
		NUS < NUSH	NUS > NUSH	NUS < NUSH	NUS > NUSH
	Q	852.508	1616.600	1866.104	1392.167
		NUS < NUSH	NUS < NUSH	NUS > NUSH	NUS < NUSH
Mathematics	ASPL	22.084	157.585	1319.786	799.288
		NUS < NUSH	NUS > NUSH	NUS > NUSH	NUS < NUSH
	CC	362.341	929.530	1490.692	691.917
		NUS < NUSH	NUS > NUSH	NUS < NUSH	NUS > NUSH
	Q	348.620	1124.274	816.541	1206.306
		NUS < NUSH	NUS < NUSH	NUS > NUSH	NUS < NUSH
Biology	ASPL	348.868	491.339	792.928	637.993
		NUS < NUSH	NUS < NUSH	NUS > NUSH	NUS < NUSH
	CC	467.957	1001.467	1411.361	276.312
		NUS < NUSH	NUS > NUSH	NUS < NUSH	NUS > NUSH
	Q	552.636	1071.332	663.916	1144.159
		NUS < NUSH	NUS < NUSH	NUS > NUSH	NUS < NUSH
Chemistry	ASPL	653.549	1765.343	1740.241	548.370
		NUS < NUSH	NUS < NUSH	NUS > NUSH	NUS < NUSH
	CC	808.564	1980.868	1480.638	3.523
		NUS < NUSH	NUS > NUSH	NUS < NUSH	NUS < NUSH
	Q	876.994	2225.229	107.692	1114.650
		NUS < NUSH	NUS < NUSH	NUS > NUSH	NUS < NUSH
Physics	ASPL	143.029	1309.520	1182.860	758.734
		NUS < NUSH	NUS < NUSH	NUS > NUSH	NUS < NUSH
	CC	347.183	1313.321	677.129	297.930
		NUS < NUSH	NUS > NUSH	NUS < NUSH	NUS > NUSH
	Q	172.122	1848.664	344.967	864.034
		NUS < NUSH	NUS < NUSH	NUS > NUSH	NUS < NUSH

We conducted Bayesian ANCOVA to compare the network measures of bootstrapped networks derived from the NUS data against the network measures of bootstrapped networks derived from the NUSH data, while also including network size a covariate in the analyses. In all comparisons, we found that the Bayes factors were well above 100[1], indicating that the network structure of the two groups indeed differed from each other, in line with the results of the frequentist approach reported in the main text. In order to improve the presentation of Table 7, we report $\log_{10}$(Bayes factors) as well as the direction of the effect below the value.

[1]Bayes factor is defined as the ratio of the likelihood of the alternative hypothesis (i.e., the network measure of the NUS group is different from the NUSH group) to the likelihood of the null hypothesis (i.e., the network measure of the NUS group is not different from the NUSH group). Based on the recommendations by Lee and Wagenmakers (2014), a Bayes factor of 100 indicates very strong evidence for the alternative hypothesis.

Table 8 Summary of abbreviations used in this paper

Network estimation methods	Network science measures	Other
community network (CN)	average shortest path length (ASPL)	Undergraduates from the National University of Singapore (NUS)
naive random walk network (NRW)	clustering coefficient (CC)	High school students from the National University of Singapore High School of Mathematics and Science (NUSH)
pathfinder network (PF)	modularity index (Q)	
correlation-based network (CR)		

Table 9 Summary of sample size-matched network estimations

Network Measure	CN		NRW		PF		CR	
	NUS	NUSH	NUS	NUSH	NUS	NUSH	NUS	NUSH
Animals								
ASPL	6.91 (1.12)***	9.145	2.98 (0.06)***	4.236	3.42 (0.17)***	2.063	3.00 (0.14)***	3.896
CC	0.29 (0.04)***	0.151	0.19 (0.01)***	0.084	0.56 (0.02)***	0.744	0.75 (0.01)***	0.707
Q	0.77 (0.03)***	0.777	0.33 (0.01)***	0.468	0.22 (0.03)***	0.063	0.60 (0.02)***	0.704
Fruits								
ASPL	3.65 (0.83)***	4.701	2.34 (0.06)***	3.429	4.82 (0.5)***	3.035	2.31 (0.06)***	2.679
CC	0.16 (0.09)***	0.121	0.41 (0.03)***	0.184	0.51 (0.03)***	0.699	0.76 (0.01)***	0.747
Q	0.51 (0.09)***	0.617	0.18 (0.01)***	0.347	0.28 (0.05)***	0.068	0.45 (0.02)***	0.514
Psychology								
ASPL	3.20 (0.88)***	2.000	4.59 (0.18)***	4.783	1.70 (0.05)***	1.601	2.43 (0.14)***	2.726
CC	0.20 (0.12)***	0.000	0.05 (0.01)***	0.051	0.82 (0.01)***	0.829	0.72 (0.02)**	0.725
Q	0.44 (0.12)***	0.219	0.56 (0.01)***	0.596	0.05 (0.01)***	0.007	0.48 (0.03)***	0.514
Mathematics								
ASPL	3.93 (1.1)***	2.156	4.34 (0.35)***	3.887	2.26 (0.1)***	1.921	2.97 (0.19)***	3.438
CC	0.27 (0.13)***	0.557	0.09 (0.01)***	0.076	0.74 (0.02)***	0.780	0.73 (0.01)***	0.708
Q	0.54 (0.11)***	0.375	0.47 (0.01)***	0.470	0.08 (0.01)***	0.019	0.58 (0.02)***	0.635
Biology								
ASPL	5.07 (1.27)**	5.209	4.42 (0.14)	4.417	1.83 (0.05)***	1.736	3.38 (0.23)	3.380
CC	0.18 (0.07)***	0.046	0.06 (0.01)***	0.056	0.78 (0.01)***	0.792	0.72 (0.01)***	0.720
Q	0.64 (0.08)***	0.653	0.54 (0.01)***	0.533	0.06 (0.01)***	0.020	0.64 (0.02)***	0.642
Chemistry								
ASPL	5.23 (1.35)***	7.047	3.86 (0.07)***	4.194	1.95 (0.06)***	1.669	3.42 (0.36)***	3.286
CC	0.24 (0.08)***	0.357	0.09 (0.01)***	0.048	0.77 (0.01)***	0.787	0.71 (0.01)***	0.723
Q	0.64 (0.07)***	0.666	0.49 (0.01)***	0.537	0.06 (0.01)***	0.030	0.63 (0.03)***	0.675
Physics								
ASPL	2.94 (0.75) ***	2.457	3.89 (0.11) ***	4.083	1.97 (0.08) ***	1.795	3.02 (0.20) ***	3.097

Table 9 continued

Network Measure	CN		NRW		PF		CR	
	NUS	NUSH	NUS	NUSH	NUS	NUSH	NUS	NUSH
CC	0.14 (0.14) ***	0.223	0.08 (0.01) ***	0.064	0.78 (0.02) ***	0.770	0.72 (0.01) ***	0.729
Q	0.41 (0.10) ***	0.321	0.47 (0.01) ***	0.489	0.08 (0.01) ***	0.019	0.59 (0.02) ***	0.616

Due to the somewhat large discrepancy in the sample sizes of NUS and NUSH group, an additional bootstrapping analysis was conducted to determine if differences in the structure of NUS and NUSH networks could have been due to differences in sample sizes. Because the NUS group was much larger than the NUSH group, our approach was to randomly select *N* number of participants from the NUS group such that *N* was equal to the number of participants in the NUSH group for each cue word. We then obtained the estimated networks for the sample-sized-matched-NUS data in the same manner as for the original set of analyses. This process was repeated 1,000 times and the descriptive statistics of the network measures of these simulated sample-size-matched-NUS networks are reported in Table 9.

One-sample *t* tests were then conducted to see if network measures of the estimated NUSH networks were significantly different from the distribution of network measures of simulated sample-size-matched-NUS networks. As Table 9 shows, almost all of these statistical comparisons are significant, giving us some confidence to say that the differences in the structure of NUS and NUSH networks are *not* merely a by-product of the larger sample sizes of the NUS group.

CN = community network; NRW = naïve random walk; PF = pathfinder; CR = correlation-based networks; NUS = National University of Singapore students; NUSH = NUS High School students; ASPL = average shortest path length; CC = clustering coefficient; Q = modularity. Values represent distribution mean with standard deviations in parentheses.

** $p < .01$, *** $p < .001$.

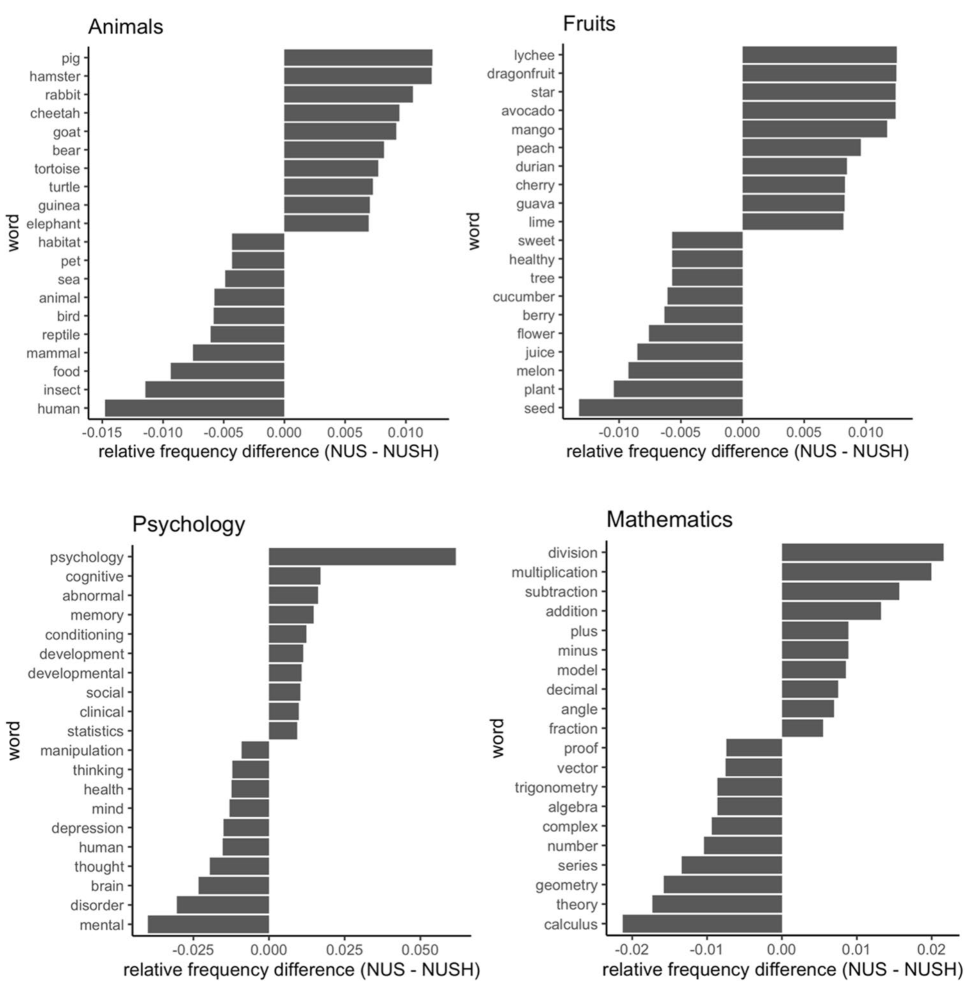

Fig. 2 Visualizations of the relative frequency of fluency responses for all cue words across NUS and NUSH groups

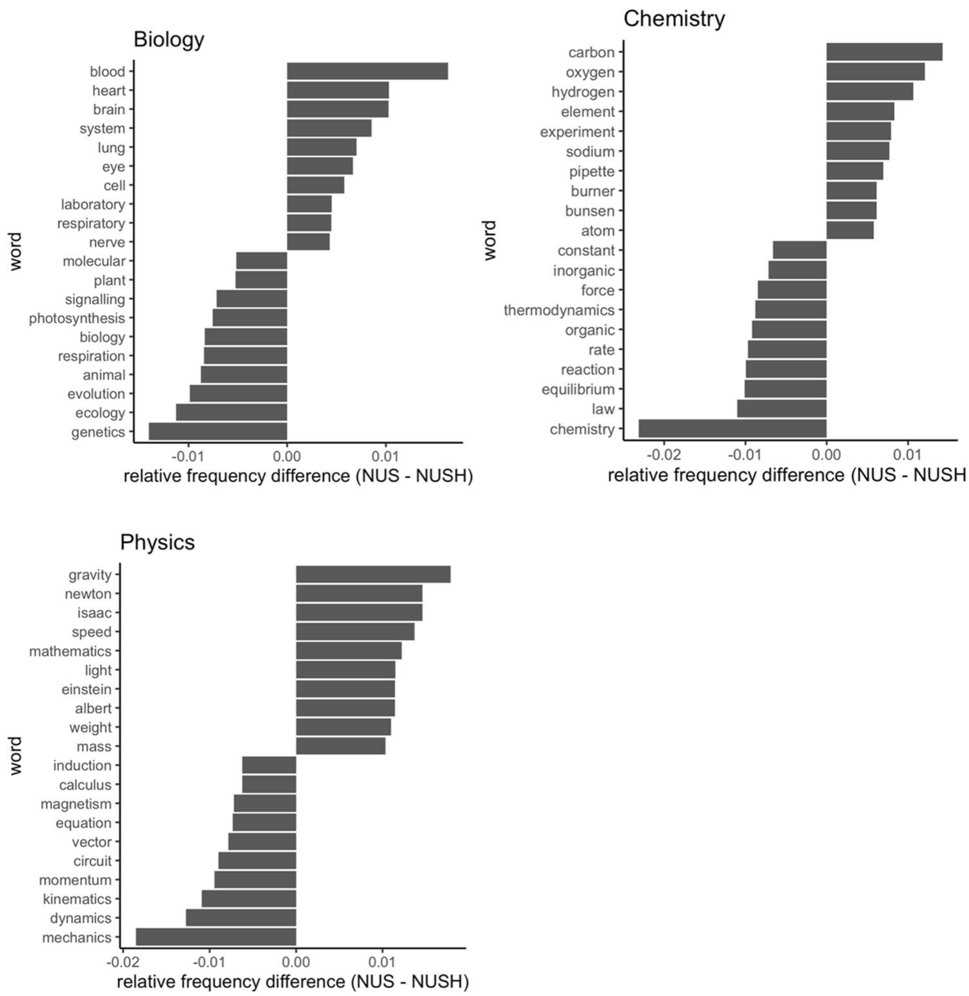

Fig. 2 (continued)

Author note This research was supported by a start-up grant (WBS R-581-000-242-133) from the National University of Singapore awarded to C.S.Q.S.

We would like to thank Dominic Cho Yu Wei, Ang Wei Boon, and Sreerajababu Aishwarya for helping to collect fluency data from NUS High students.

All data, analysis scripts, and anonymized fluency lists can be found online (https://osf.io/yjzma/).

References

Amunts, J., Camilleri, J. A., Eickhoff, S. B., Heim, S., & Weis, S. (2020). Executive functions predict verbal fluency scores in healthy participants. *Scientific Reports, 10*(1), 11141. https://doi.org/10.1038/s41598-020-65525-9

Anderson, J. R., & Bower, G. H. (1980). *Human associative memory: A brief edition*. Erlbaum.

Baer, J. (1998). The case for domain specificity of creativity. *Creativity Research Journal, 11*(2), 173–177. https://doi.org/10.1207/s15326934crj1102_7

Baer, J. (2015). The Importance of Domain-Specific Expertise in Creativity. *Roeper Review, 37*(3), 165–178. https://doi.org/10.1080/02783193.2015.1047480

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLOS ONE, 6*(5), Article e19348.

Beckage, N. M., & Colunga, E. (2019). Network growth modeling to capture individual lexical learning. *Complexity, 2019*, 1–17. https://doi.org/10.1155/2019/7690869

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics, 2008*(10), P10008–P10012. https://doi.org/10.1088/1742-5468/2008/10/P10008

Borodkin, K., Kenett, Y. N., Faust, M., & Mashal, N. (2016). When pumpkin is closer to onion than to squash: The structure of the second language lexicon. *Cognition, 156*, 60–70.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

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.

Chi, M. T. H. (2006). Laboratory Methods for Assessing Experts' and Novices' Knowledge. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. E. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 167–184). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.010

Christensen, A. P., Kenett, Y. N., Cotter, K. N., Beaty, R. E., Silvia, P. J., & Mõttus, R. (2018). Remotely close associations: Openness to experience and semantic memory structure. *European Journal of Personality, 32*(4), 480–492. https://doi.org/10.1002/per.2157

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407–428.

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods, 51*, 987–1006.

De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods, 40*(1), 213–231.

Disessa, A. A., & Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education, 20*(10), 1155–1191.

Driver, R., & Erickson, G. (1983). Theories-in-action: Some theoretical and empirical issues in the study of students' conceptual frameworks in science. *Studies in Science Education, 10*(1), 37–60. https://doi.org/10.1080/03057268308559904

Dubossarsky, H., De Deyne, S., & Hills, T. T. (2017). Quantifying the structure of free association networks across the lifespan. *Developmental Psychology, 53*(8), 1560–1570.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*(1), 1–26. https://doi.org/10.1214/aos/1176344552

Fortunato, S. (2010). Community detection in graphs. *Physics Reports, 486*(3), 75–174.

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

Goñi, J., Arrondo, G., Sepulcre, J., Martincorena, I., Vélez de Mendizábal, N., Corominas-Murtra, B., Bejarano, B., Ardanza-Trevijano, S., Peraita, H., Wall, D. P., & Villoslada, P. (2011). The semantic organization of the animal category: Evidence from semantic verbal fluency and network theory. *Cognitive Processing, 12*(2), 183–196. https://doi.org/10.1007/s10339-010-0372-x

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review, 119*(2), 431.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks. *Psychological Science, 20*(6), 729–739. https://doi.org/10.1111/j.1467-9280.2009.02365.x

Hills, T. T., Todd, P. M., & Goldstone, R. L. (2008). Search in external and internal spaces: Evidence for generalized cognitive search processes. *Psychological Science, 19*(8), 802–808. https://doi.org/10.1111/j.1467-9280.2008.02160.x

Jones, M. N., & Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114*(1), 1–37.

Kenett, Y. N., Anaki, D., & Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Frontiers in Human Neuroscience, 8*, 407.

Kenett, Y. N., Beaty, R. E., Silvia, P. J., Anaki, D., & Faust, M. (2016). Structure and flexibility: Investigating the relation between the structure of the mental lexicon, fluid intelligence, and creative achievement. *Psychology of Aesthetics, Creativity, and the Arts, 10*(4), 377–388. https://doi.org/10.1037/aca0000056

Kenett, Y. N., Wechsler-Kashi, D., Kenett, D. Y., Schwartz, R. G., Ben-Jacob, E., & Faust, M. (2013). Semantic organization in children with cochlear implants: Computational analysis of verbal fluency. *Frontiers in Psychology, 4*, 543. https://doi.org/10.3389/fpsyg.2013.00543

Kinchin, I. M., Hay, D. B., & Adams, A. (2000). How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational Research, 42*(1), 43–57.

Kleinberg, J. M. (2000). Navigation in a small world. *Nature, 406*(6798), 845.

Koponen, I. T., & Nousiainen, M. (2014). Concept networks in learning: Finding key concepts in learners' representations of the interlinked structure of scientific knowledge. *Journal of Complex Networks, 2*(2), 187–202.

Koponen, I. T., & Pehkonen, M. (2010). Coherent knowledge structures of physics represented as concept networks in teacher education. *Science & Education, 19*(3), 259–282.

Landauer, T. K. (2007). *Handbook of latent semantic analysis*. Erlbaum.

Lerner, A. J., Ogrocki, P. K., & Thomas, P. J. (2009). Network graph analysis of category fluency testing. *Cognitive and Behavioral Neurology, 22*(1), 45–52. https://doi.org/10.1097/WNN.0b013e318192ccaf

Lonie, J. A., Herrmann, L. L., Tierney, K. M., Donaghey, C., O'Carroll, R., Lee, A., & Ebmeier, K. P. (2009). Lexical and semantic fluency discrepancy scores in aMCI and early Alzheimer's disease. *Journal of Neuropsychology, 3*(1), 79–92. https://doi.org/10.1348/174866408X289935

Lydon-Staley, D. M., Zhou, D., Blevins, A. S., Zurn, P., & Bassett, D. S. (2021). Hunters, busybodies and the knowledge network building associated with deprivation curiosity. *Nature Human Behaviour, 5*(3), 327–336. https://doi.org/10.1038/s41562-020-00985-7

Morais, A. S., Olsson, H., & Schooler, L. J. (2013). Mapping the structure of semantic memory. *Cognitive Science, 37*(1), 125–145. https://doi.org/10.1111/cogs.12013

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research, 76*(3), 413–448.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences, 103*(23), 8577–8582.

Persky, A. M., & Robinson, J. D. (2017). Moving from novice to expertise and its implications for instruction. *American Journal of Pharmaceutical Education, 81*(9), 6065–6080. https://doi.org/10.5688/ajpe6065

Reder, L. M., Park, H., & Kieffaber, P. D. (2009). Memory systems do not divide on consciousness: Reinterpreting memory in terms of activation and binding. *Psychological Bulletin, 135*(1), 23–49. https://doi.org/10.1037/a0013974

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255.

Salkowski, L. R., & Russ. (2018). Cognitive processing differences of experts and novices when correlating anatomy and cross-sectional imaging. *Journal of Medical Imaging, 5*(3), 1–17. https://doi.org/10.1117/1.JMI.5.3.031411

Schvaneveldt, R. W. (1990). *Pathfinder associative networks: Studies in knowledge organizations*. Ablex.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology, 5*, 772. https://doi.org/10.3389/fpsyg.2014.00772

Siew, C. S. Q. (2018). Using network science to analyze concept maps of psychology undergraduates. *Applied Cognitive Psychology, 33*(4), 662–668.

Siew, C. S. Q. (2020). Applications of network science to education research: Quantifying knowledge and the development of expertise through network analysis. *Education Sciences, 10*(4), 101. https://doi.org/10.3390/educsci10040101

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81*(3), 214–241.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29*(1), 41–78. https://doi.org/10.1207/s15516709cog2901_3

Troyer, A. K., Moscovitch, M., Winocur, G., Alexander, M. P., & Stuss, D. O. N. (1998). Clustering and switching on verbal fluency: The effects of focal frontal-and temporal-lobe lesions. *Neuropsychologia, 36*(6), 499–504.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic Press.

van Wijk, B. C. M., Stam, C. J., & Daffertshofer, A. (2010). Comparing brain networks of different size and connectivity density using graph theory. *PLOS ONE, 5*(10), e13701. https://doi.org/10.1371/journal.pone.0013701

Walker, P., & Finney, N. (1999). Skill development and critical thinking in higher education. *Teaching in Higher Education, 4*(4), 531–547. https://doi.org/10.1080/1356251990040409

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440–442.

Whitaker, M. M., Pointon, G. D., Tarampi, M. R., & Rand, K. M. (2020). Expertise effects on the perceptual and cognitive tasks of indoor rock climbing. *Memory & Cognition, 48*(3), 494–510. https://doi.org/10.3758/s13421-019-00985-7

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal fluency: Language or executive function measure? *Applied Neuropsychology. Adult, 23*(1), 29–34. https://doi.org/10.1080/23279095.2015.1004574

Wolff, C. E., Jarodzka, H., & Boshuizen, H. P. A. (2017). See and tell: Differences between expert and novice teachers' interpretations of problematic classroom management events. *Teaching and Teacher Education, 66*, 295–308. https://doi.org/10.1016/j.tate.2017.04.015

Wulff, D. U., De Deyne, S., Jones, M. N., Mata, R., & Consortium, A. L. (2019). New perspectives on the aging lexicon. *Trends in Cognitive Sciences, 23*(8), 686–698.

Zemla, J. C., & Austerweil, J. L. (2018). Estimating semantic networks of groups and individuals from fluency data. *Computational Brain & Behavior, 1*(1), 36–58.

Christensen, A. P., & Kenett, Y. (2019). *Semantic Network Analysis (SemNA): A tutorial on preprocessing, estimating, and analyzing semantic networks*. *PsyArXiv.* https://doi.org/10.31234/osf.io/eht87

Jun, K.-S., Zhu, J., Rogers, T. T., Yang, Z., & others. (2015). Human memory search as initial-visit emitting random walk. *Advances in Neural Information Processing Systems*, *28*, 1072–1080.

Lee, M., & Wagenmakers, E. (2014). *Bayesian cognitive modeling: A practical course*. *Cambridge University Press.* https://doi.org/10.1017/CBO9781139087759

Linn, M. C. (2006). *The knowledge integration perspective on learning and instruction.* Cambridge University Press.

Novak, J. D. (2010). *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations*. *Routledge.*

Siew, C. S. Q., Wulff, D. U., Beckage, N. M., & Kenett, Y. N. (2019, *2019*). Cognitive Network Science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity*, 2108423.

Wulff, D. U., De Deyne, S., Aeschbach, S., & Mata, R. (2021). *Understanding the aging lexicon by linking individuals' experience, semantic networks, and cognitive performance*. PsyArXiv. https://doi.org/10.31234/osf.io/z3ebt

Zemla, J. C., Cao, K., Mueller, K. D., & Austerweil, J. L. (2020). SNAFU: The Semantic Network and Fluency Utility. *Behavior Research Methods*, 1–19.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.