New Ideas in Psychology 71 (2023) 101034

Contents lists available at ScienceDirect

New Ideas in Psychology

journal homepage: www.elsevier.com/locate/newideapsych

Hypergraph models of the mental lexicon capture greater information than pairwise networks for predicting language learning

Salvatore Citraro [a,1], Judy Warner-Willich [b,1], Federico Battiston [c], Cynthia S.Q. Siew [d], Giulio Rossetti [a], Massimo Stella [e,*]

[a] Institute of Information Science and Technologies “A. Faedo” (ISTI), National Research Council (CNR), G. Moruzzi 1, Pisa, Italy
[b] Department of Computer Science, University of Exeter, Exeter, UK
[c] Department of Network and Data Science, Central European University, Vienna, Austria
[d] Department of Psychology, National University of Singapore, Singapore
[e] Department of Psychology and Cognitive Science, University of Trento, Trento, Italy

ARTICLE INFO

Keywords:
Cognitive modelling
Network science
Language acquisition
Cognitive hypergraphs

ABSTRACT

Human memory is a complex system that works in associative ways: Reading a cue word can lead to the recollection of associated concepts. The network structure of memory recall patterns has been shown to contain insights about a wide variety of cognitive phenomena, including language acquisition. However, most current network approaches use pairwise connections, i.e. links between only two words at a time. This ignores the possibility that more than two concept representations might be simultaneously associated in memory. We overcome this modelling limitation by introducing *cognitive hypergraphs* as models of human memory. We model memory recall patterns through word associations from the *Small World of Words* project for *N=6003* concepts (Study 1) and for *N=497* concepts (Study 2). In each study we represent word associations as either a pairwise network or a hypergraph. By combining psycholinguistic norms and network centrality measures with machine learning, we quantitatively investigate whether there is any benefit to using the hypergraph model over a pairwise network in predicting test-based age of acquisition norms in children up to age 9 years (Study 1) or normative learning in toddlers up to age 30 months (Study 2, based on CHILDES data). We show that cognitive hypergraphs capture more information than pairwise networks from the same data: Cognitive hypergraphs are considerably more powerful than pairwise networks at predicting age of acquisition trends in toddlers, children and teenagers. Our studies showcase how novel approaches merging artificial intelligence and higher-order interactions can help us understand cognitive development.

1. Introduction

Language learning implies building conceptual representations of words in the human mind (Skehan, 1998). These representations constitute a cognitive dictionary, a mental lexicon where knowledge is stored and processed (Aitchison, 2012). Unlike brain tissues, this cognitive system cannot be directly accessed nor tinkered with within a lab setting. Exploring the structure of associations between conceptual representations requires the mediation of knowledge-related tasks that involve the mental lexicon (Aitchison, 2012; Castro & Siew, 2020). Free associations (Nelson et al. 2000, 2003; Wilson & Kiss, 1988) represent a cognitive task that is becoming increasingly popular in: (i) investigating the associative organisation of concepts in the mental lexicon (De Deyne & Storms, 2008; Steyvers & Tenenbaum, 2005) and, (ii) exploring the interplay between the structure of conceptual associations and cognitive phenomena like language learning (Hills et al., 2009) and processing (De Deyne et al., 2013). Specifically, free associations represent memory recall patterns from the mental lexicon and can be gathered through empirical recall tasks (Brysbaert et al., 2000; Stella et al., 2019; Wilson & Kiss, 1988). In the continuous free association task introduced by (De Deyne et al., 2013), an individual reads a cue (e.g. “pen”) and is asked to produce the first 3 associations that come to their mind (e.g. “letter”, “paper”, “writing”). These associations do not explicitly follow specific requirements (e.g. consider only associations that are synonyms) and

* Corresponding author.
E-mail address: massimo.stella@inbox.com (M. Stella).
[1] These authors contributed equally.

https://doi.org/10.1016/j.newideapsych.2023.101034
Received 18 November 2022; Received in revised form 22 May 2023; Accepted 23 May 2023
Available online 11 June 2023

represent the outcome of recall processes from human memory, which can be modelled in different ways, e.g. through a signal originating on the cue and spreading through its conceptual associates (see Nelson et al., 2003). Free associations can map memory recall patterns from the cognitive knowledge available to an individual (Wulff et al., 2022). Several works indicate that free associations can map how conceptual representations are associated inside human memory, providing a network proxy, that is, a model approximating the structure of associative knowledge, cf. (Kumar, 2021).

Modelling free associations as a network leads to a representation of memory where nodes represent concepts and links indicate their recall from memory (De Deyne & Storms, 2008; Steyvers & Tenenbaum, 2005). Recent advancements at the interface of psychology and network science have highlighted the benefit of modelling free associations as a cognitive construct related to a variety of cognitive phenomena (Kenett et al., 2018). The availability of large-scale datasets for free associations, like the Edinburgh Associative Thesaurus (Wilson & Kiss, 1988), the Florida Free Association Norms (Nelson et al., 2004) and, more recently, the Small World of Words (De Deyne et al., 2019), have enabled over 20 years of data-informed studies about cognitive modelling with free associations. Networks of free associations/memory recall patterns can shed light on a wide range of cognitive phenomena (Wulff et al., 2022). The number of free associations of individual words was shown to be predictive of early language learning in toddlers (Hills et al., 2009; Stella et al., 2017; Citraro et al., 2023). More free associations also corresponded to facilitative patterns in lexical identification tasks in healthy adults (i.e. distinguishing words from non-words in the shortest time possible) (De Deyne et al., 2013). The smallest number of free associations separating any two concepts was found to be predictive of semantic relatedness, outperforming other state-of-art models in psychology like latent semantic analysis (Kenett et al., 2017; Kumar et al., 2020).

All of the above studies used pairwise networks to model free association data, i.e. networks where links can connect only two nodes at a time (Newman, 2018). Pioneering work by De Deyne and colleagues (De Deyne et al., 2013) showed that across several lexical identification tasks, pairwise networks considering only links between a cue and its associations outperformed other pairwise ways of linking free associations in predicting several lexical processing tasks. However, cognitive models that permit only two items to interact with each other at a time can be quite limiting. In the continuous task, up to three items are recalled after reading a cue and this process might indicate interactions (potentially with different strengths) taking place between all concepts rather than between couples of them (De Deyne et al. 2013, 2019). In continued free association tasks, past investigations reported considerable evidence that first responses are more semantically related to cues, whereas such similarity decreases for second and third responses (Bruza et al., 2009; Nelson et al., 2003). This difference motivated the creation of models for serial recall, like Entanglement and Spooky-action-at-a-distance, including differences in the ways first and subsequent responses are generated (see for more details Nelson et al., 2003 and Galea et al., 2011). Although these approaches discuss signals spreading according to concept similarity and not based on a more specific network structure, they are relevant from a network perspective, since they highlight a limitation of pairwise network models for free association data. The latter builds links between a cue and all its responses, however evidence from Galea et al. (2011) indicates that the link between a cue and its first response should be qualitatively different from other links (between the same cue and subsequent responses). Not building links between pairs of words while still clustering them as a single recollection might address such a limitation.

This work aims to introduce and investigate cognitive hypergraphs as a novel idea in psychology to cluster recollections together and thus model the structure of memory recall patterns from the mental lexicon and compare it against pairwise networks to predict language learning in children/teenagers (Study 1) and toddlers (Study 2) within a machine learning framework.

1.1. Related literature

Network science is the field concerned with understanding a system by analysing the relations between its components (Newman, 2018; Siew et al., 2019). In any network, nodes/vertices represent the individual units of the system being represented (e.g. concepts in memory) and are connected by edges/links, representing interactions between units (e.g. memory recalls). Complex networks were extensively adopted to model human memory already in the 1970s, see also the review by (Castro & Siew, 2020). Networks were used to describe data storage hierarchies in early computers that could be applied to modelling human memory, see for instance (Quillian, 1967). The pioneering work by Collins and Loftus reviewed these early approaches and suggested a cognitive model for lexical retrieval based on a network of semantic feature sharing between concepts, e.g. “cat” and “dog” sharing the feature “having a tail” and thus being linked (Collins & Loftus, 1975). However, the authors did not test how their model performed on empirical data, and complex networks gained more attention in psychology only thirty years later (Siew, 2020; Siew et al., 2019). Steyvers and Tenenbaum built three semantic networks, including one of free associations with more than 6000 English words (Steyvers & Tenenbaum, 2005). Importantly, Steyvers and Tenenbaum showed how pairwise networks of free associations could become computational frameworks for understanding language learning dynamics. The authors showed how the degree of words in these networks, i.e. the number of links of a word, could explain variance in empirical age of acquisition ratings. These results were replicated on a larger dataset by (De Deyne & Storms, 2008), who also found significant negative associations between degree and self-reported age of acquisition ratings: Words reported to be acquired earlier tended to have more free associations. Analogous results were also found in early language learning, as measured in toddlers between 18 and 30 months of age (Hills et al., 2009; Stella et al., 2017): Degree in free associations was found to be more predictive of the normative age of acquisition ranking of words compared to degree in other networks made of synonyms (where concepts are linked if overlapping in meaning) or of phonological similarities (where concepts are linked if differing by the addition/substitution/deletion of one phoneme). A similar result held also for closeness centrality, i.e. the inverse average network distance between a node and all the others connected to it (Newman, 2018). How could one interpret these patterns from a psychological perspective? Degree and closeness centrality capture, respectively, the number of links and the mental shortcuts (i.e., shortest paths) a node is involved in. Having more associations and being linked to other concepts through fewer associations are measures of prominence within memory (Kumar, 2021). Higher degree/closeness corresponding to earlier acquisition might thus reflect a tendency for humans to learn more prominent concepts in memory first, a cognitive strategy known as preferential acquisition (Hills et al., 2009; Siew & Vitevitch, 2020).

The above works relied consistently on pairwise networks. Pairwise links involve only two nodes and this can be a modelling limitation when considering the mental lexicon: Mental search processes might include multiple categories of conceptual similarities at once (Aitchison, 2012; Hills & Kenett, 2022; Kumar, 2021), so that the recalls should be considered as a single entity (e.g. {“pen”, “letter”, “paper”, “writing”}, where “pen” is the cue and the latter 3 concepts are responses to the cue) rather than as distinct links (e.g. {“pen”, “letter”}, {“pen”, “paper”}, {“paper”, “writing”}). Hypergraphs can account for these types of structures. Hypergraphs are complex networks where multiple nodes can co-exist within the same link, called a “hyperedge” (Battiston et al., 2020). Introduced in the 1970s (Berge, 1973), hypergraph models received considerable attention in quantitative modelling only within the last few years, mainly thanks to the growth of statistical models of hypergraph construction and analysis (Battiston et al., 2020; Young et al., 2021; Musciotto et al., 2021). For a clear and extensive review of hypergraph models, we refer the interested reader to the work of

Battiston and colleagues (Battiston et al., 2020). Noticeably, their review contains a gap: applications of hypergraph models in psychology and the cognitive sciences.

Pairwise networks disregard relationships of conceptual similarities between recalls in free association data because such relationships did not provide significant information across explaining variance in several lexical tasks (cf. De Deyne et al., 2013). Going beyond pairwise networks, hypergraphs can encode conceptual similarities between serially recalled responses within a single hyperlink, i.e. a generalisation of a link that connects more than 2 concepts at once. Encapsulating cues and responses within a single hyperlink provides a representation of memory recall patterns going beyond cue-response relationships, as encoded in pairwise links, and including conceptual similarities due to memory recall processes (the weak conceptual relationships present between responses). In contrast to pairwise networks, hyperlinks encode information about which groups of responses were associated to a given cue: in pairwise networks, once links are built, there is no information about how cues and responses clustered together, and whereas such knowledge about concept clustering, reflecting mental search processes (Hills & Kenett, 2022), is indeed preserved by hyperlinks. This crucial difference makes hypergraphs different from pairwise networks and potentially more suitable for modelling how conceptual similarities can influence memory recall and other cognitive patterns.

Imposing a pairwise network, where associations are encoded by links connecting two items at a time, might be a modelling assumption unable to account for the simultaneous presence of multiple categories of associations in human memory. As discussed in (De Deyne et al., 2013), continued free associations constitute a serial recall of items from memory. Recent research showed that serial recall from memory follows a foraging strategy (cf. Hills & Kenett, 2022): mental search identifies candidates for recall mostly in a cluster of concepts sharing several semantic features; however, after the search exhausted all such candidates (i.e. after foraging depletes a patch), a transition to a different cluster is performed. Consequently, serial recall will produce groups of items that are conceptually related, albeit weakly (De Deyne et al., 2019; Kumar, 2021).

Hypergraphs are mathematical objects where hyperedges encode interactions between any number of nodes at a time (Battiston et al., 2020; Berge, 1973), and have recently proven to be a useful tool for the analysis of many real-world relational systems across a variety of domains. Indeed, higher-order (non-pairwise, multiple) interactions were found to be ubiquitous in a wide range of systems, including group formation in face-to-face human (Cencetti et al., 2021) and animal (Musciotto et al., 2022) social networks, collaboration systems (Patania et al., 2017), ecosystems (Grilli, Barabás, Michalska-Smith, & Allesina, 2017) and even the human brain (Santoro et al., 2022). Moreover, taking into account the presence of such non-pairwise interactions has allowed us to better understand the function and behaviours of different systems (Battiston et al., 2021), from the diffusion of information (Iacopini et al., 2019) to the emergence of human cooperation (Alvarez-Rodriguez et al., 2021) in networked populations. Surprisingly, hypergraphs and their potential to properly describe non-pairwise associations have been largely neglected in the case of cognitive systems. Operationally, cognitive hypergraphs allow to represent free associations among an arbitrary number of items as hyperedges, without the need to decide for any specific pairwise network construction (e.g. build links only between a cue and its associations).

1.2. Outline of paper

This paper aims to introduce and test hypergraph networks of free associations in predicting age of acquisition norms. Through machine learning, two studies will investigate the potential benefit of hypergraphs over pairwise networks for predicting either continuous (Brysbaert & Biemiller, 2017) or ordinal (Stella et al., 2017) age of acquisition (AoA) data. Both a pairwise network and a hypergraph will be constructed using free associations from the Small World of Words (De Deyne et al., 2019). Degree and closeness centrality of individual concepts in these networks will be used as features for a machine learning regression of 6000 continuous AoA norms (Study 1) and for a rank prediction task of 500 ordinal AoA norms (Study 2). These aspects are described in the Methods section. Using model performance assessment we will test which representation of human memory provides the better sets of features for interpreting AoA data (Results). Our findings and the importance of individual features will be discussed with respect to the relevant psychology literature (Discussion), and finally we introduce cognitive hypergraphs as an innovative modelling framework of cognitive processes (Conclusions).

2. Materials and methods

In this section we describe the datasets and the methods used to build the pairwise/hypergraph networks and set up the machine learning section. Data and methods are different across the two studies. The first study uses word association data from (De Deyne et al., 2019) to predict continuous words' age of acquisition from (Brysbaert & Biemiller, 2017). The second study uses the same word association data to predict ordinal age of acquisition as ranked by the CHILDES Project dataset (MacWhinney, 2000). Pairwise networks and hypergraph representations of the same word association dataset are used to extract words' centralities as features for machine learning.

2.1. The network science of pairwise networks and hypergraphs

This section serves as an introduction for the formalism used throughout Study 1 and 2.

Graph notions. A graph $G = (V, E)$ is composed of a set V of nodes and a set E of edges which are unordered pairs of elements of V (Newman, 2018). The connections between nodes exist on a pairwise basis, i.e. an edge connects exactly two nodes together. An example of a graph with six nodes is shown in Fig. 1 (left). In our formalism, we will use "graph", "pairwise network" and "network" as synonyms. There are various measures estimating how 'central' a node is within a given graph/network. Two of the most common ones are degree and closeness centrality (Newman, 2018). Degree centrality measures how well connected a certain node is by counting its links. In Fig. 1 (left), node *f* has a degree of 1. To understand closeness centrality, we first define a *path* between nodes *i* and *j* as a sequence of links connecting them. In any path, nodes can be visited only once. Nodes for which there is at least one path connecting them are said to be connected. Then, the closeness centrality of a node *i* is defined as:

$$C(i) = \frac{N-1}{\sum_j d(i,j)}, \qquad \text{Eq. 1}$$

where N is the number of nodes connected to i and $d(i,j)$ is the smallest number of links connecting i and j, i.e. the network distance (Newman, 2018). Hence, closeness centrality equals the inverse average network distance between one node and all its connected neighbours. The higher the closeness $C(i)$, the fewer links that have to be traversed to reach any connected neighbour when starting from i. In Fig. 1 (left), node *f* has the

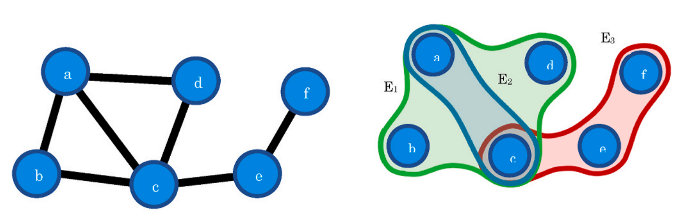

Fig. 1. A pairwise network (left) and a hypergraph (right), both with 6 nodes.

lowest closeness centrality as it is the furthest apart from all other nodes.

Hypergraph notions. A hypergraph $H = (V_H, E_H)$ is composed of a set V_H of nodes and a set E_H of non-empty subsets of V called hyperedges. Hypergraphs expand on the traditional graph structure by allowing for higher-order interactions between more than two units by means of hyperedges. A hypergraph with all hyperedges of size 2 is equivalent to a pairwise network/graph. The hypergraph in Fig. 1 (right) has 6 nodes connected by 3 hyperedges of sizes 4, 3 and 2. Indeed, the clusters of nodes, i.e. the hyperlinks, in Fig. 1 are reminiscent of community structure, where nodes in a pairwise network are clustered together. As mentioned in the Introduction, a key difference between hypergraphs and pairwise networks is that the first preserves information about clustering between responses and cues, i.e. how many responses were provided together to a given cue.

We can define degree and closeness centrality on hypergraphs as well (Battiston et al., 2020). Node degree in hypergraph is still a count of how many hyperedges contain a given node. Distance-based computations, like closeness, on hypergraphs can be made less computationally demanding by recasting hyperedges into a bipartite data structure (Aksoy et al., 2020). We leverage on the vertex-based notion of the *s-framework* for hypergraphs (Aksoy et al., 2020) by computing Equation (1) on the node-projection of a hypergraph. Any hypergraph can be represented as a bipartite data structure such that V_H and E_H are two distinct sets and the elements of V_H are connected to the elements of E_H according to how nodes are sorted in hyperedges. We want to compute the closeness of elements of V_H so we can focus on them through a node-projection, i.e. simplifying the bipartite structure by considering only nodes and linking them if they share at least s connections with elements of V_H in the bipartite data structure. In other words, (i) the node-projection of the hypergraph is a data structure where nodes are connected if they share at least s hyperedges; (ii) the closeness formula provided in Equation (1) is applied on this different structure, where the s-prefix highlights the width/strength of connectivity between two nodes. With s = 1, node f is connected to nodes e and c in the node-projection of the hypergraph in Fig. 1 (right), because the three nodes share the hyperedge E3. Thus, f's 1-closeness is higher than its corresponding closeness in the graph structure. With s = 2, however, node f loses all its connections, not allowing computing f's 2-closeness. The parameter s can be thought of as the width of connection between two nodes, assuming that nodes sharing more hyperedges are tighter than nodes sharing less hyperedges. The parameter s can be tuned between the minimum and maximum sizes of hyperedges, thus leading to a set of closeness scores for a single node (e.g. one for s = 1, one for s = 2 and one for s = 3). Closeness centralities will indicate how closely connected one node is to its neighbours when different higher-order mechanisms are considered (Aksoy et al., 2020).

Evaluation of machine learning models. For model assessment we use the following measures (Alpaydin, 2020): the Kendall's tau correlation coefficient (τ), the Pearson's correlation coefficient (ρ), the Root Mean Squared Error (RMSE), the Coefficient of Determination (R^2), and the normalised Discounted Cumulative Gain (NDCG). These values estimate a model's ability at predicting the desired output (Alpaydin, 2020). Kendall's tau correlation coefficient measures the degree of agreement between two columns of ranked data, with a value of 0 representing no correlation and a value of 1 representing a perfect correlation. Pearson's correlation coefficient is another statistical measure of the strength of relationship between two variables that measures how close the data points are to a line of best fit. It can take any value in the range [−1,1] with −1 representing a perfect negative correlation, +1 a perfect positive correlation, and 0 no correlation. The root mean squared error (RMSE) is a measure of error which is calculated as the square root of the averaged sum of squares of the error between predicted values and the true values. The coefficient of determination (R^2) is a regression score, with a value of 1 being the best. It measures the proportion of variance that is predicted by the regressor. Finally, the normalised discounted cumulative gain (NDCG) is a measure of rank strength that gives the higher ranked values a greater score. It can take any value in the range [0,1], with 1 being a perfect score.

2.2. Network data and construction

Data description and cleaning. We use word association data from the *Small World of Words* (SWoW) project (De Deyne et al., 2019). SWoW is an online large-scale database modelling multilingual mental lexicons through a memory recall game, where participants are asked to respond with at most 3 words that come to mind given a cue word (e.g. what do you think of when reading "letter"?). In Study 1 and 2 we use the English SWoW dataset, with more than 1.2 million responses structured as lists {"cue", "association_1", "association_2", "association_3"}, e.g. {"pen", "letter", "writing", "ink"}.

Association data are complemented by words' age of acquisition (AoA) norms. In Study 1, we use the continued age of acquisition values from (Brysbaert & Biemiller, 2017), obtained by directly asking adults about their knowledge of words at various ages. We focus on early language learning, considering words acquired before age 9 years. This threshold was selected based on results by (Stella et al., 2018), who showed a drastic shift in word learning strategies based on network structure after that age. In Study 2, we use ordinal age of acquisition values as ranked by the CHILDES Project dataset (MacWhinney, 2000) and previously used in (Stella et al., 2017). Normative word learning is represented as a ranking: Words in top positions were found to be produced by more children, in child-adult conversations, at an earlier age in months. This dataset focuses on normative early language learning between 18 and 30 months. We use the words in these two datasets to filter out the SWoW data and reduce the computational costs for building networks and computing node centralities. Hence, data filtering is contextual to the performed task: Of the 13000+ cue words in the SWoW data, 6003 were used in Study 1, while 497 were used in Study 2.

Network construction. We build the pairwise and the higher-order network representations by leveraging the word association data. For the pairwise network construction, word associations are processed into a list of pairwise connections, where the cue word was connected to each of the subsequent user-inputted responses. Thus, the cue and response words are connected as in Fig. 2 (a), i.e., as in the *R123* procedure used in the original reference paper (De Deyne et al., 2019). For instance, if participants' data associate the cue "book" with the responses "fairy" and "horror", then the links "book" - "fairy" and "book" - "horror" are built. De Deyne and colleagues showed that this approach led to networks with the highest explanatory values for several psycholinguistic tasks (see De Deyne et al., 2019 for more details). For the hypergraph construction, for each recall list, cue and responses are connected by means of a single hyperedge, as in the toy example of Fig. 2 (b). For instance, with reference to the above example, a single hyperlink of size 3 would be built, i.e. {"book", "fairy", "horrow"}. Notice how the size of the hyperlink reflects the number of associates provided for a given cue. Once built, links in a pairwise network are always size 2 (i.e. they include two nodes). Instead, hyperlinks built from free association data can have size 2, 3 or 4, according to how many responses were given to a given cue. An example of the real hypergraph instead is shown in Fig. 2 (c). We have $|V| = 6003$ and $|E| = 229505$ for the pairwise network and $|V_H| = 6003$ and $|E_H| = 207915$ in Study 1. For Study 2, we have $|V| = 497$ - $|E| = 3112$ for the pairwise network and $|V_H| = 497$ and $|E_H| = 2649$ for the hypergraph. The NetworkX Python package (https://networkx.org/, Accessed: 02/02/2022) was used to build the pairwise graph models of word associations, and HyperNetX for the hypergraph models (https://pnnl.github.io/HyperNetX/build/index.html, Accessed: 02/02/2022). Notice that in the underlying free association data, only 90.5% of concepts were featured as both cues and responses, meaning that roughly 1 in 10 concepts were only considered as a response. Not being given the chance for a given concept to be featured as a cue creates some sampling biases that make it difficult to consider directionality as psychologically relevant information. For this reason, we considered undirected pairwise

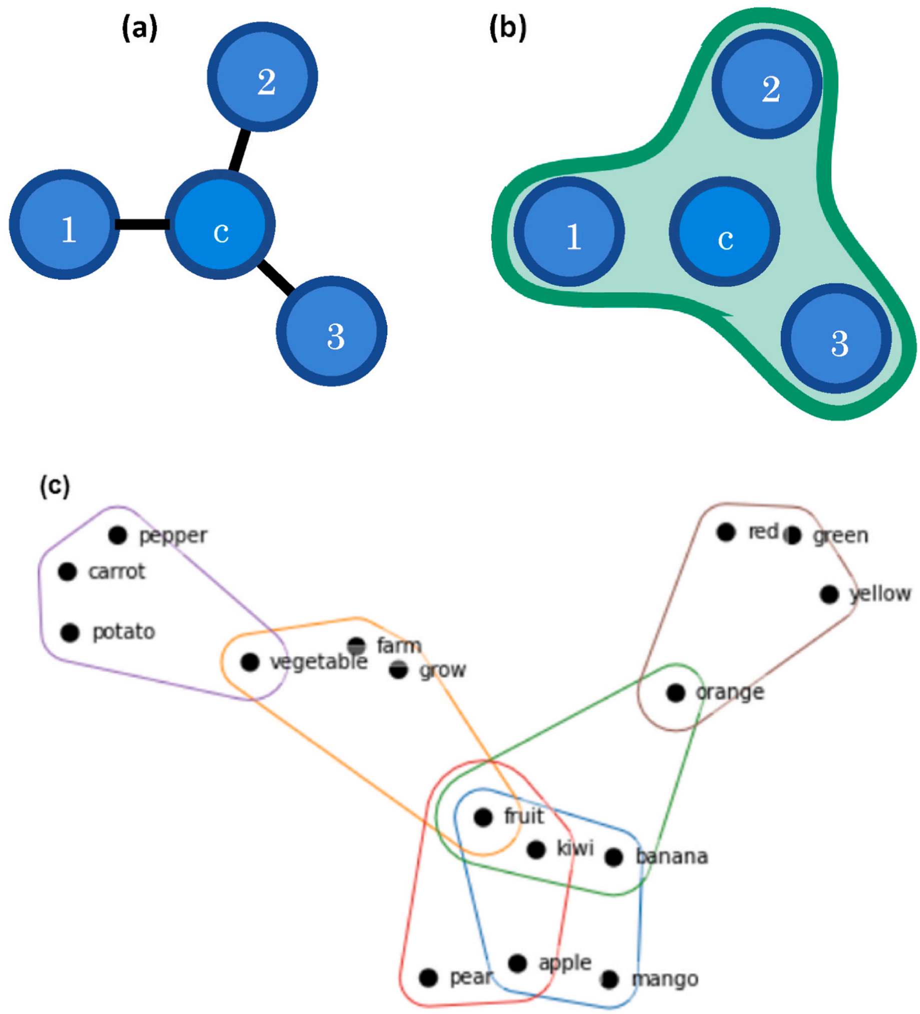

Fig. 2. Graph (a) and hypergraph (b) interaction construction, and an example of a hypergraph piece based on the association data (c).

networks - as in previous approaches (Hills et al., 2009; Kenett et al., 2018; Stella et al., 2019) - and undirected hypergraphs.

Machine learning setting. Logistic regression, support vector machines and random forests (Alpaydin, 2020) were tested in Study 1 but only random forests provided R^2 higher than 0 when trained on the features reported in the following. All models tested here were implemented in *scikit-learn* 1.2.1 in Python. The random forest regressor was trained using only one network centrality feature (i.e. node degree) and two psycholinguistic features (word frequency and length). Previous works on normative learning in children older than 30 months found that degree and closeness provided equivalent results and were redundant (Stella et al. 2017, 2018). Hence, in Study 1 we used degree as the only network feature of the regressor. The logarithm of word frequency and word length were found to be the most important features in regression tasks for the AoA norms at hand (Brysbaert & Biemiller, 2017), and were thus included as features. Other psycholinguistic norms like polysemy (Stella, 2019) were left out for the sake of model simplicity. In general terms, random forests partition data points according to their features to construct an ensemble of decision trees (Alpaydin, 2020). Such an ensemble, or forest, is then adopted to create multiple estimates which are rather robust to noise. The regressor was tuned to use 150 trees and a mean squared error criterion for measuring the quality of a split. The aim of this regressor was to reproduce the numerical values of the AoA norms by (Brysbaert & Biemiller, 2017).

For Study 2, we used an XGBoost ranker to reproduce the ranking in which concepts are acquired by most children over time. Over a given ranking, XGBoost works to produce an optimal ranking objective function (Alpaydin, 2020), which learns how to reproduce a given ranking of data points (concepts in our case) through a set of numerical features (e.g. network centralities of concepts). The aim of this ranking method was to reproduce the normative learning order/ranking in which most children in the CHILDES dataset learned concepts over time. XGBoost was configured to use 150 estimators and a learning rate of 0.2, selected by computing the error rate with a varied learning rate and choosing a value that gives the smallest error. The ranker is trained using degree and s-closeness centrality (s=*1,2,3*). Previous work found that word frequency and length provided very limited insights for reproducing the normative age of acquisition normative ranking (Stella et al., 2017). Hence, for the sake of model simplicity, we neglected these features and concentrated only on degree and closeness, which previously predicted the normative ranking to different extents within very early language learning (Stella et al., 2017).

The different features adopted in Study 1 and Study 2 make the studies unsuitable for direct comparison, due to the different features adopted in each study which were selected based on previous work. For these reasons, we focus on comparing hypergraphs and pairwise networks within each study.

To improve robustness to noise, model results are obtained through a Monte Carlo process (Alpaydin, 2020), where results are averaged over 100 independent trainings of the prediction models, each relative to random splits of training/testing data (80% training/20% testing). Moreover, to analyse the effect that each feature (e.g. network degree, frequency, length or closeness centrality) has on the performance of the model, i.e. to estimate feature importance (Alpaydin, 2020), we performed feature reshuffling: for each feature, the values of the feature's column in the data were shuffled and the machine learning models retrained with the shuffled values. Then, predictions were made after the shuffling, and measures calculated. This process was repeated 100 times to obtain average values of the change in measures for each shuffling (cf. Results). Reshuffling and re-training aims to quantitatively test how much model performance degrades when each individual feature is substituted with noise (having the same mean and variance, thanks to the reshuffling process). For instance, by reshuffling node degree and comparing model performance against the results from the original model, trained on non-shuffled features, we can quantify how much degree contributes to the regression/ranking. This reshuffling thus implements feature importance analysis (see Alpaydin, 2020). Results are reported as mean values presented together with standard errors.

For model assessment, we consider the Root Mean Squared Error (RMSE), Kendall Tau, Pearson's r coefficient and the R^2 measure. We also count the percentage P_ε of results that are correctly predicted below a tolerance ε. In mathematical terms, let us denote with d_i a datapoint, including the features for concept i. The machine learning model maps d_i into an AoA prediction $\hat{y}_i$ or an AoA ranking prediction $\hat{Y}_i$ for concept i. The RMSE can then be expressed as:

$$RMSE = \sqrt{\frac{\sum_i^N (\hat{y}_i - y_i)^2}{N}}$$

where N is the number of concepts available in the model and y_i is the empirical AoA value. The same formula, interpreted as a measurement of dispersion, holds for the predicted $\hat{Y}_i$ and the empirical value. Kendall Tau and Pearson's r coefficient are both correlation measures (see Alpaydin, 2020), although the latter is based on ranking agreement and can thus be easily interpreted when comparing rankings (as we do in Study 2). The R^2 measure quantifies how much variance observed in the empirical AoA values is explained by model prediction, for more details see (Alpaydin, 2020). We define P_ε as:

$$P_\varepsilon = \frac{\#\ of\ concepts\ i\ such\ that\ |\hat{y}_i - y_i| < \varepsilon}{N},$$

where the numerator counts the number of concepts whose predicted AoA norms (or rankings) differ from its empirical value by less than a quantity ε (which thus measures years, for Study 1, or rankings, for Study 2). P_ε thus depends on ε and can give rise to profiles where different models can be compared (as we do in Figs. 3 and 4).

3. Results

3.1. Study 1: continuous age of acquisition regression

Table 1 and Table 2 show the results for the pairwise graph and the hypergraph models, respectively. The values (columns 'Original') highlight the benefit of the hypergraph representation against the graph. The higher-order representation performs better than the pairwise network model, e.g., the hypergraph gives a 6% lower error rate and a Kendall tau correlation value 8% higher – cf. also Fig. 3 (left). In addition, Fig. 3 (right) shows the percentage of the predicted results that were within a margin 'epsilon' of the empirical value. The hypergraph model can be seen to have a slight edge, with 59.4% of the predictions within ±1 year of the true value, whereas the graph model predicted 54.7% correctly within the same range.

In this study, degree is the only network feature used for training. Tables 1 and 2 highlight the performances of the two network models with respect to each feature shuffling. The shuffling of degree has the greatest impact in both cases, with a 6.3% increase in root mean squared error on the graph model and a 12.7% increase on the hypergraph model. Hence, network degree is the most important feature for the model.

The shuffling of word length and frequency have a minor effect. For instance, to verify if this was due to randomness or not, a Mann-Whitney U test conducted on these results shows that the change in error from word length shuffling on both models is significant at the 5% level ($U=1705$, $p=0.0315$ for the hypergraph; $U=3301$, $p=0.00086$ for the pairwise network). Therefore, only word length does have a small effect on the prediction. A similarly reduced effect for frequency and length over these AoA norms was found also in (Brysbaert & Biemiller, 2017).

3.2. Study 2: ordinal age of acquisition ranking

Similar to the previous section, Table 3 and Table 4 show the results for the pairwise graph and the hypergraph models respectively. The values (columns 'Original') highlight the benefit of the hypergraph representation against the graph, with a root mean squared error value 36% lower than that from the graph model and a much higher R^2 value of 0.6656 versus only 0.1807 from the graph model – cf. also Fig. 4 (left). In addition, Fig. 4 (right) shows the percentage of the predicted results that were within a margin 'epsilon' of the empirical value. The hypergraph makes 77.1% of the predictions within ±50 positions of the empirical rank, whereas the graph model predicts only 36.6% in the same range.

These results provide strong quantitative evidence that the hypergraph provides more information that correlates with the ordering of word acquisition, through the use of degree and s-closeness centrality. Notice that the scores for different values of s-closeness (s = 1,2,3) were

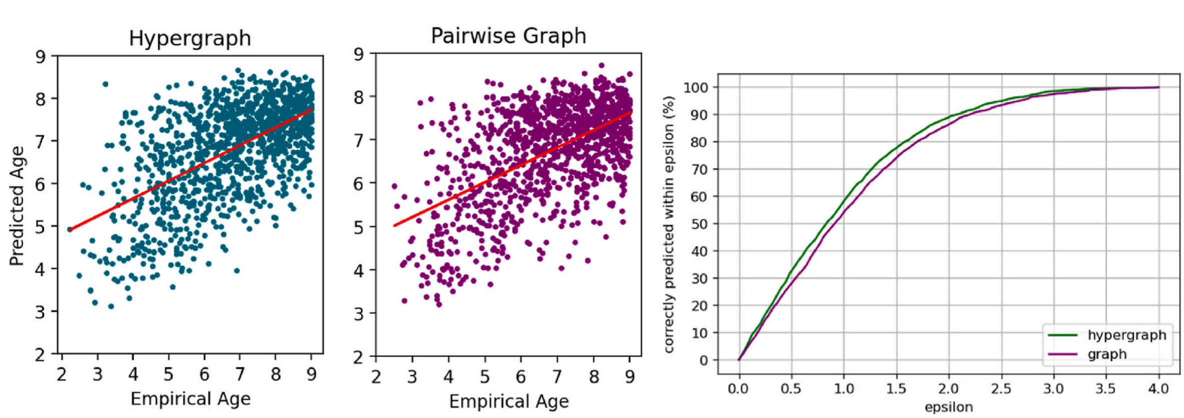

Fig. 3. The relationship between predicted and empirical age for hypergraph and graph models (left) and the percentage of results that are correctly predicted within ±epsilon years of the empirical value (right).

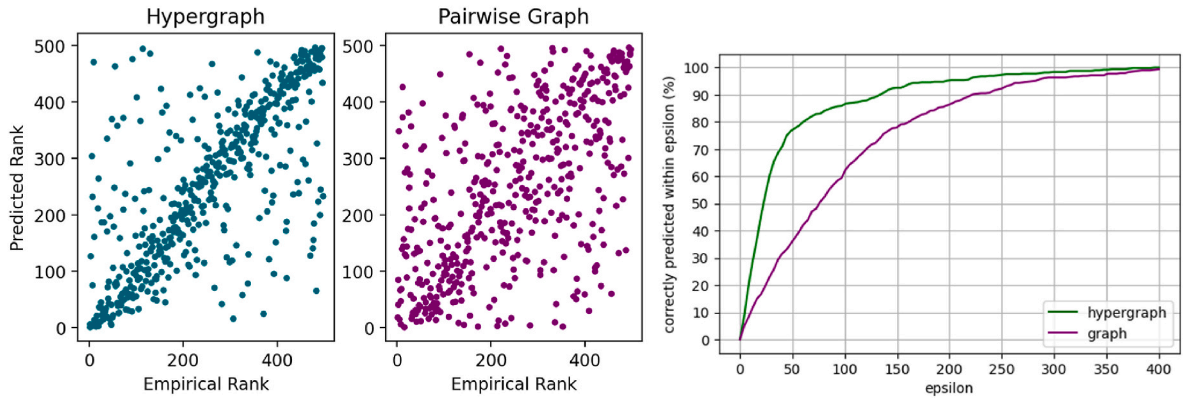

Fig. 4. The relationship between predicted and empirical age rank for hypergraph and graph models (left) and the percentage of results that are correctly predicted within ±epsilon rankings of the empirical value (right).

Table 1
For a pairwise network model, performance and feature importance results for the regression task. Model performance is based in terms of root mean squared and correlations between model's estimates and empirical AoA norms. The largest absolute difference for each measure is highlighted in bold. Relative to the original model, reshuffling node degree degrades performance the most compared to frequency and length.

Measure	Original Performance	Degree Shuffled		Logarithmic Frequency Shuffled		Word Length Shuffled	
		Value	Difference	Value	Difference	Value	Difference
Root Mean Squared Error	1.3217	1.4045	**+0.0828**	1.3503	+0.0286	1.3416	+0.0199
Kendall Tau	0.3181	0.2593	**−0.0588**	0.3072	−0.0110	0.3135	−0.0046
Pearson's r	0.5285	0.4293	**−0.0992**	0.5006	−0.0279	0.5199	−0.0086
R^2	0.2662	0.1680	**−0.0982**	0.2318	−0.0344	0.2434	−0.0228

Table 2
For a hypergraph model, performance and feature importance results for the regression task. Model performance is based in terms of root mean squared and correlations between model's estimates and empirical AoA norms. The largest absolute difference for each measure is highlighted in bold. Relative to the original model, reshuffling node degree degrades performance the most compared to frequency and length.

Measure	Original Performance	Degree Shuffled		Logarithmic Frequency Shuffled		Word Length Shuffled	
		Value	Difference	Value	Difference	Value	Difference
Root Mean Squared Error	1.2440	1.4014	**+0.1574**	1.2742	+0.0302	1.2521	+0.0081
Kendall Tau	0.3832	0.2671	**−0.1161**	0.3557	−0.0275	0.3671	−0.0161
Pearson's r	0.6057	0.4467	**−0.1590**	0.5729	−0.0328	0.5878	−0.0179
R^2	0.3521	0.1676	**−0.1845**	0.3171	−0.0350	0.3354	−0.0167

Table 3
For a pairwise network model, performance and feature importance results for the regression task. Model performance is based in terms of root mean squared and correlations between model's estimates and the empirical AoA ranking. The largest absolute difference for each measure is highlighted in bold. Relative to the original model, reshuffling closeness degrades performance the most compared to degree.

Measure	Original Performance	Degree Shuffled		Closeness Centrality Shuffled	
		Value	Difference	Value	Difference
Root Mean Squared Error	129.86	143.62	+13.76	187.33	**+57.47**
Kendall Tau	0.4254	0.3513	−0.0741	0.0992	**−0.3262**
Pearson's r	0.5904	0.4990	−0.0914	0.1475	**−0.4429**
R^2	0.1807	−0.0020	−0.1827	−0.7049	**−0.8856**
NDCG	0.9589	0.9507	−0.0082	0.9010	**−0.0579**

found to be different for less than 5% of the data points and were thus averaged over s. In other words, we considered averages of s-centrality for different values of s. Notice these averages displayed relative error margins - standard errors over mean - lower than 0.01%. Hence, in this study, the model used a single average s-closeness value rather than

Table 4
For a hypergraph model, performance and feature importance results for the regression task. Model performance is based in terms of root mean squared and correlations between model's estimates and the empirical AoA ranking. The largest absolute difference for each measure is highlighted in bold.

Measure	Original Performance	Degree Shuffled		s-Closeness Centrality Shuffled	
		Value	Difference	Value	Difference
Root Mean Squared Error	82.96	81.03	−1.93	85.78	**+2.82**
Kendall Tau	0.6969	0.7012	**+0.0043**	0.6987	+0.0018
Pearson's r	0.8328	0.8405	+0.0077	0.8212	**−0.0116**
R^2	0.6656	0.6710	+0.0054	0.6425	**−0.0231**
NDCG	0.9824	0.9890	**+0.0066**	0.9864	+0.0040

3 different ones.

Tables 3 and 4 highlight the performances of the two network models with respect to each feature shuffling. For instance, in the graph-based model, closeness centrality has the largest effect on prediction quality in the graph model, with a 44.3% increase in RMSE versus only a 10.6% increase when the degree column was shuffled. Interestingly, the column shuffling on the hypergraph model shows some apparent increases in performance. This is likely due to many features being used and so there is redundancy in the model for loss of data and therefore a change in one column does not affect the model's prediction ability. Nonetheless, the closeness centrality shuffling appears to give the largest reduction in prediction quality. A similarly prominent relevance for closeness in predicting normative learning was found in previous work (Stella, 2019; Stella et al., 2017).

4. Discussion

Our results quantitatively highlight a benefit of the hypergraph model over the pairwise network model in identifying patterns of early language learning from free association data. Even through the sole use of degree on the hypergraph, the regression model was able to predict AoA better than on the pairwise graph, with a 6% lower error rate. However, the model overall did not make accurate predictions for AoA. This is likely due to the lack of ability to calculate closeness centrality, as degree is not enough on its own to accurately predict the values. Nonetheless, the results are promising and work as a proof of concept of the use of the hypergraph in this setting. Additionally, the ranking task provided further evidence for the benefit of hypergraphs with its performance, predicting over two times as many results in the range of 50 ranks of the correct value and providing a lower error rate of 36%. Hence, it seems that the hypergraph is able to capture more information about the structure of conceptual similarities in regard to age of acquisition. This research provides promise for the future use of hypergraphs in this setting.

Given that the hypergraph has been shown to be a better model than the pairwise network model, at least within the context of word acquisition, what is the implication of such a result for human learning and the nature of mental representations?

First, consider that the computational resources required to construct and analyse hypergraph models were very high, which forced us to limit the number of word associations that could be utilised in the hypergraph. In the context of human memory, we may also expect that the representation of multiple, higher-order clusters of simultaneously interacting and co-activated concepts in memory is also cognitively quite expensive relative to the simpler pairwise memory representation of concepts.

Does it make sense for a cognitive system to be investing high levels of cognitive resources in order to construct and maintain such a complex memory representation of concepts? Below we consider the extant literature on the development of children's semantic representations and offer a tentative explanation for the superiority of hypergraphs, and why it is plausible for human memory to have a hypergraph organisation despite its complexity.

The development of categorization is an important component of human cognition. Specifically, categorization refers to the ability to classify entities as members of the same category despite differences among those entities. For instance, understanding that "cat" and "dog" are examples of the "animal" category. Although developmental psychologists debate on the specific ways that children develop this ability, it is generally accepted that a crucial starting point is that children have the prerequisite perceptual and attentional mechanisms that enables them to extract multiple correspondences among features shared by category members (Goldstone, 1994; Sloutsky, 2003; Sloutsky & Fisher, 2004). In particular, basic level categories (e.g., animals, birds, food) have correlated features (e.g., most birds can fly, all birds have feathers and wings) that a human learner must be able to uncover and consolidate over the course of development (Sloutsky, 2003).

From the perspective of a hypergraph representation of the lexicon, it is plausible to reimagine the development of hyperedges among sets of concepts as a by-product of the extraction of higher-order regularities in the learner's environments (e.g., robin, pigeon, feather, wing, etc. could be concepts that are connected by a hyperedge in the hypergraph).

When considering the overall trajectory of the development of a child's semantic knowledge, research in the developmental sciences generally indicates that the development of associative relations between concepts (e.g., spoon-soup) occurs earlier than taxonomic relations (e.g., pasta-rice; Unger & Fisher, 2021; Sloutsky et al., 2017). Associative relations can be readily acquired through exposure to direct co-occurrences of concepts and words in the language environment (e.g., cup-water), whereas the acquisition of taxonomic relations bootstrap the shared co-occurrence of concepts (e.g., cup-water and cup-juice leads to the taxonomic relation between water-juice as types of liquids that can be found in a sip cup; Sloutsky et al., 2017).

The acquisition of taxonomic relations appears to be more gradual than associative relations among children (Unger et al., 2020). If we consider the acquisition of these relations within the context of a network model, it is clear that while pairwise associations among two concepts may characterise a very early lexicon, the nature of semantic knowledge organisation demands more sophisticated representations of higher order semantic relationships among concepts.

The observation that the hypergraph model was a better predictor of age of acquisition norms than the pairwise network model suggests that the ability to specify hyperedges that connect sets of related concepts that are larger than two may be especially important in quantifying the acquisition of higher order semantic relations (with taxonomic relations as one example) in the developing mental lexicon.

Given the above, we ultimately suggest that hypergraphs depict cognitive "compartmentalizations" that are lingering vestiges of learning in early childhood and perhaps more generally of human learning. This suggestion could also explain why the hypergraph model may approximate human memory better and show better prediction of age of acquisition data as compared to the pairwise network. Furthermore, there is not necessarily an additional computational/cognitive cost to constructing these higher-order representations; rather, such representations emerge naturally as natural sets of concepts that are learned early on in life and continue to remain "grouped" or linked together into adulthood.

To sum up, our findings provide quantitative evidence that hypergraphs are models of relevance for investigating cognitive phenomena. Although past approaches in computational psychology introduced concepts similar to hyperlinks, e.g. synsets as collections of related synonyms in WordNet (Miller, 1995), the psychological literature has mostly re-cast such structures into pairwise networks, i.e. see the pairwise network investigation by Sigman and Cecchi (2002). Hypergraphs as networks of hyperlinks have been discussed in recent works (Hills & Kenett, 2022) but their potential in quantitative, comparative investigation has yet to be explored. We believe our approach provides rather encouraging evidence for the power and usefulness of hypergraphs as a novel idea in data-powered psychological models.

4.1. Limitations and Future directions

On a final note, it is important for us to clarify the limitations of our modelling approach and highlight how improvements can be made in the future. Firstly, the free association data used here was gathered from adults but used to predict early language learning in children. This choice was motivated by other past approaches showing that even free association data produced by adults provide significant insights in predicting the variance observed in AoA norms for children (see Hills et al., 2009; Stella et al., 2017; Steyvers and Tenenbaum 2005). Within the ever growing Small World of Words project (De Deyne et al., 2019) it would be interesting to see if free association data provided by children (when available in the future) can lead to better results through the framework outlined here. Secondly, not all the data was used in constructing the graphs. Due to computational limitations, the SWoW dataset had to be reduced significantly so that calculations could be made in a suitable time frame. Despite using a limited subset of the data, the results were still conclusive of the relative superiority of the hypergraph over the pairwise network.

In modelling terms, our adopted hypergraph model considered hyperlinks of varying size between 2 and 4 but it represents only one among a wide category of hypergraph models, e.g. directed hypergraphs (where hyperlinks have an inner directionality), k-uniform hypergraphs (where all hyperlinks include k nodes) and many others (see Battiston et al., 2021). Despite the sheer number of different models, in our exploratory study we had to select a single model and were guided by the cognitive interpretation of free associations as serial recall tasks of clusters of concepts relative to a given cue. Future investigations might delve more into assessing the impact of considering or excluding the cue from hyperlinks. However, excluding cues would void comparison with pairwise network models, where cues engage in links with responses. We consider this an exciting research opportunity for network scientists interested in cognitive science and psychology.

Another limitation of this study is its focus on normative learning, i.e. how most children learn words over time. Given the promising results obtained here by cognitive hypergraphs, future investigations should test whether these cognitive models are more predictive than pairwise networks even when considering how individual children learn words over time (see the approach of Beckage & Colunga, 2019). These investigations of individual lexical learning would require more fine grained association data, which might come from initiatives similar to the recent MySWOW proof-of-concept project (Wulff et al., 2022), which gathered free associations for 8 individuals. Other types of network data might be plugged into the model, e.g. sharing semantic features might better capture the hierarchical nature of semantic memory, influencing model performance. The main difficulty of performing such steps would be about how to interpret hyperlinks: In free association data, one hyperlink maps all the outcomes of a single mental search process, i.e. a cue and all its responses, and this leads to a clearer interpretation of the hypergraph building process.

In the future, we expect to see advances also in the algorithms used to construct and analyse hypergraphs, which currently require considerable computational costs. Furthermore, access to powerful computing infrastructure will enable larger hypergraphs to be built. With the use of supercomputers, the potential for vast amounts of data gathered from hypergraph structures may be possible in future work. When combined with the fast growing numbers of rich datasets of linguistic and behavioural norms, we anticipate that the hypergraph modelling approach will enable cognitive scientists to build network representations that better reflect the sophistication and complexity of the human lexicon.

Author contributions

Conceptualization: JWW and MS; Data curation: JWW, MS and SC; Formal analysis: JWW and SC; Investigation: All authors; Methodology: JWW, SC and MS; Supervision: MS; Validation: CSQ, FB and GR; Visualization: JWW and SC; Roles/Writing - original draft: All authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon*. John Wiley & Sons.

Aksoy, S. G., Joslyn, C., Marrero, C. O., Praggastis, B., & Purvine, E. (2020). Hypernetwork science via high-order hypergraph walks. *EPJ Data Science, 9*(1), 16.

Alpaydin, E. (2020). *Introduction to machine learning*. Cambridge, USA: MIT press.

Alvarez-Rodriguez, U., Battiston, F., de Arruda, G. F., Moreno, Y., Perc, M., & Latora, V. (2021). Evolutionary dynamics of higher-order interactions in social networks. *Nature Human Behaviour, 5*(5), 586–595.

Battiston, F., Amico, E., Barrat, A., Bianconi, G., Ferraz de Arruda, G., Franceschiello, B., … Petri, G. (2021). The physics of higher-order interactions in complex systems. *Nature Physics, 17*(10), 1093–1098.

Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., … Petri, G. (2020). Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports, 874*, 1–92.

Beckage, N. M., & Colunga, E. (2019). Network growth modeling to capture individual lexical learning. *Complexity*, 2019.

Berge, C. (1973). *Graphs and hypergraphs*. Amsterdam, The Netherlands: North-Holland Publishing.

Bruza, P., Kitto, K., Nelson, D., & McEvoy, C. (2009). Extracting spooky-activation-at-a-distance from considerations of entanglement. In *Quantum interaction: Third international symposium, QI 2009, Saarbrücken, Germany, March 25-27, 2009. Proceedings* (Vol. 3, pp. 71–83). Springer Berlin Heidelberg.

Brysbaert, M., & Biemiller, A. (2017). Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Methods, 49*(4), 1520–1523.

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica, 104*(2), 215–226.

Castro, N., & Siew, C. S. (2020). Contributions of modern network science to the cognitive sciences: Revisiting research spirals of representation and process. *Proceedings of the Royal Society A, 476*(2238), Article 20190825.

Cencetti, G., Battiston, F., Lepri, B., & Karsai, M. (2021). Temporal properties of higher-order interactions in social networks. *Scientific Reports, 11*(1), 1–10.

Citraro, S., Vitevitch, M. S., Stella, M., & Rossetti, G. (2023). Feature-rich multiplex lexical networks reveal mental strategies of early language learning. *Scientific Reports, 13*(1), Article 1474.

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

De Deyne, S., Navarro, D. J., Perfors, A., et al. (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., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods, 45*(2), 480–498.

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

Galea, D., Bruza, P., Kitto, K., Nelson, D., & McEvoy, C. (2011). Modelling the activation of words in human memory: The spreading activation, spooky-activation-at-a-distance and the entanglement models compared. In *Quantum interaction: 5th international symposium, QI 2011, Aberdeen, UK, June 26-29, 2011, revised selected papers 5* (pp. 149–160). Springer Berlin Heidelberg.

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition, 52*(2), 125–157. https://doi.org/10.1016/0010-0277(94)90065-5

Grilli, J., Barabás, G., Michalska-Smith, M. J., & Allesina, S. (2017). Higher-order interactions stabilize dynamics in competitive network models. *Nature, 548*(7666), 210–213.

Hills, T. T., & Kenett, Y. N. (2022). Is the mind a network? Maps, vehicles, and skyhooks in cognitive network science. *Topics in Cognitive Science, 14*(1), 189–208.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science, 20*(6), 729–739.

Iacopini, I., Petri, G., Barrat, A., & Latora, V. (2019). Simplicial models of social contagion. *Nature Communications, 10*(1), 1–9.

Kenett, Y. N., Levi, E., Anaki, D., & Faust, M. (2017). The semantic distance task: Quantifying semantic distance with semantic network path length. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(9), 1470.

Kenett, Y. N., Levy, O., Kenett, D. Y., Stanley, H. E., Faust, M., & Havlin, S. (2018). Flexibility of thought in high creative individuals represented by percolation analysis. *Proceedings of the National Academy of Sciences, 115*(5), 867–872.

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review, 28*(1), 40–80.

Kumar, A. A., Balota, D. A., & Steyvers, M. (2020). Distant connectivity and multiple-step priming in large-scale semantic networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(12), 2261.

MacWhinney, B. (2000). *The CHILDES project: The Database* (Vol. 2). Psychology Press.

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM, 38*(11), 39–41.

Musciotto, F., Battiston, F., & Mantegna, R. N. (2021). Detecting informative higher-order interactions in statistically validated hypergraphs. *Communications Physics, 4*(1), 1–9.

Musciotto, F., Papageorgiou, D., Battiston, F., & Farine, D. R. (2022). *Beyond the dyad: Uncovering higher-order structure within cohesive animal groups*. bioRxiv.

Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory & Cognition, 28*(6), 887–899.

Nelson, D. L., McEvoy, C. L., & Pointer, L. (2003). Spreading activation or spooky action at a distance? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(1), 42.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402–407.

Newman, M. (2018). *Networks*. London, UK: Oxford University Press.

Patania, A., Petri, G., & Vaccarino, F. (2017). The shape of collaborations. *EPJ Data Science, 6*, 1–16.

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science, 12*(5), 410–430.

Santoro, A., Battiston, F., Petri, G., & Amico, E. (2022). *Unveiling the higher-order organization of multivariate time series. arXiv preprint arXiv:2203.10702*.

Siew, C. S. (2020). Feature distinctiveness effects in language acquisition and lexical processing: Insights from megastudies. *Cognitive Processing, 21*(4), 669–685.

Siew, C. S., & Vitevitch, M. S. (2020). An investigation of network growth principles in the phonological language network. *Journal of Experimental Psychology: General, 149* (12), 2376.

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

Sigman, M., & Cecchi, G. A. (2002). Global organization of the Wordnet lexicon. *Proceedings of the National Academy of Sciences, 99*(3), 1742–1747.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences, 7*(6), 246–251. https://doi.org/10.1016/S1364-6613(03)00109-8

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in Young children: A similarity-based model. *Journal of Experimental Psychology: General, 133*(2), 166–188. https://doi.org/10.1037/0096-3445.133.2.166

Sloutsky, V. M., Yim, H., Yao, X., & Dennis, S. (2017). An associative account of the development of word learning. *Cognitive Psychology, 97*, 1–30. https://doi.org/10.1016/j.cogpsych.2017.06.001

Stella, M. (2019). Modelling early word acquisition through multiplex lexical networks and machine learning. *Big Data and Cognitive Computing, 3*(1), 10.

Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. *Scientific Reports, 7*(1), 1–10.

Stella, M., Beckage, N. M., Brede, M., & De Domenico, M. (2018). Multiplex model of mental lexicon reveals explosive learning in humans. *Scientific Reports, 8*(1), 1–11.

Stella, M., De Nigris, S., Aloric, A., & Siew, C. S. (2019). Forma mentis networks quantify crucial differences in STEM perception between students and experts. *PLoS One, 14* (10), Article e0222870.

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.

Unger, L., & Fisher, A. V. (2021). The emergence of richly organized semantic knowledge from simple statistics: A synthetic review. *Developmental Review, 60*, Article 100949. https://doi.org/10.1016/j.dr.2021.100949

Unger, L., Savic, O., & Sloutsky, V. M. (2020). Statistical regularities shape semantic organization throughout development. *Cognition, 198*, Article 104190. https://doi.org/10.1016/j.cognition.2020.104190

Wilson, M., & Kiss, G. (1988). *EAT: The Edinburgh associative corpus*. Oxford Text Archive Core Collection.

Wulff, D. U., De Deyne, S., Aeschbach, S., & Mata, R. (2022). Using network science to understand the aging lexicon: Linking individuals' experience, semantic networks, and cognitive performance. *Topics in Cognitive Science, 14*(1), 93–110.

Young, J. G., Petri, G., & Peixoto, T. P. (2021). Hypergraph reconstruction from network data. *Communications Physics, 4*(1), 1–11.