EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE

Open Access

Towards hypergraph cognitive networks as feature-rich models of knowledge

Salvatore Citraro[1]* , Simon De Deyne[2], Massimo Stella[3†] and Giulio Rossetti[1†]

*Correspondence:
salvatore.citraro@isti.cnr.it
[1]Institute of Information Science
and Technologies "A. Faedo" (ISTI),
National Research Council (CNR), Via
G. Moruzzi, 1, Pisa, Italy
Full list of author information is
available at the end of the article
†Equal contributors

Abstract

Conceptual associations influence how human memory is structured: Cognitive research indicates that similar concepts tend to be recalled one after another. Semantic network accounts provide a useful tool to understand how related concepts are retrieved from memory. However, most current network approaches use pairwise links to represent memory recall patterns (e.g. reading "airplane" makes one think of "air" and "pollution", and this is represented by links "airplane"-"air" and "airplane"-"pollution"). Pairwise connections neglect higher-order associations, i.e. relationships between more than two concepts at a time. These higher-order interactions might covariate with (and thus contain information about) how similar concepts are along psycholinguistic dimensions like arousal, valence, familiarity, gender and others. We overcome these limits by introducing feature-rich cognitive hypergraphs as quantitative models of human memory where: (i) concepts recalled together can all engage in hyperlinks involving also more than two concepts at once (cognitive hypergraph aspect), and (ii) each concept is endowed with a vector of psycholinguistic features (feature-rich aspect). We build hypergraphs from word association data and use evaluation methods from machine learning features to predict concept concreteness. Since concepts with similar concreteness tend to cluster together in human memory, we expect to be able to leverage this structure. Using word association data from the Small World of Words dataset, we compared a pairwise network and a hypergraph with $N = 3586$ concepts/nodes. Interpretable artificial intelligence models trained on (1) psycholinguistic features only, (2) pairwise-based feature aggregations, and on (3) hypergraph-based aggregations show significant differences between pairwise and hypergraph links. Specifically, our results show that higher-order and feature-rich hypergraph models contain richer information than pairwise networks leading to improved prediction of word concreteness. The relation with previous studies about conceptual clustering and compartmentalisation in associative knowledge and human memory are discussed.

Keywords: Cognitive networks; Free associations; Feature-rich networks; Hypergraphs

1 Introduction

Words in language bear implicit, unexpressed features [1]. When reading "the pen is on the table", we immediately consider "pen" as a concrete object, even though the sentence does not convey specific quantitative information about it [2]. We think of "building" as

Springer

something with a large size, of "love" as something abstract, of "crime" as something negative [3]. These features contribute to making human language complex and nuanced just as much as its cognitive reflection in the human mind [1]. Theoretical models [1, 4–6] informed by considerable experimental evidence [7–10] point out that linguistic knowledge is organised in an associative way, with ideas sharing many features being more tightly connected and easier to be acquired, processed and recalled one after another. The cognitive system apt at processing knowledge expressible with language is commonly called "mental lexicon" [1, 10]. Differently from common dictionaries, the mental lexicon includes not only knowledge relative to meanings but also other phonological [5], emotional [11] and visual [12] aspects of conceptual knowledge, among many other features [2, 3].

Quantitative investigations of the mental lexicon, its structure and functioning, have recently benefited from the advent of Big Data and network science [6, 13]. Massive psycholinguistic experiments mapped thousands of concepts across multiple dimensions, providing quantitative estimates for word concreteness, imageability, valence/sentiment, arousal and many other features (cf. [3]). Access to this big data fostered the creation of several large-scale network models, with thousands of nodes, representing knowledge in the mental lexicon as engaging in different types of conceptual associations [13–15]. Feature-sharing networks (which are different from feature-rich networks [16]) link concepts based on overlap in semantic features [17], or overlap in sounds, in the case of phonological networks [5], or concept similarity in the case of synonymy networks [17], among many other possibilities. The proliferation of network representations, backed up by psychological theory, saw even more refined attempts at directly mapping memory recall patterns from the mental lexicon: Free associations map cue-target responses from memory, devoid of any specific semantic or phonological constraint affecting them [15, 18, 19]. Reading the prompt/cue "book" and immediately thinking of "chapter" creates a free association link "book" – "chapter". Continued free associations extend this task to consider up to three recalls [15], e.g. reading "math" elicits "bad", "hard", "wrong" in an individual [20]. Modelling continued free associations as three cue-response links led to the creation of free association networks better suited to capture weak associations compared to single-response procedures as [15, 18]).

From a knowledge modelling perspective, free association networks have been a valid approach to capture semantic cognition broadly, as previous work demonstrates they capture semantic relatedness between concepts [21], differentiate individuals based on creativity [22], reflect the affective (positive/negative) connotations of concepts [20, 23]. This has a range of applications as well. For example, word associations can be used to infer psychometric measures of mental distress in healthy populations [24]. The Small World of Words is a multilingual international research project on free associations, gathering millions of free associations across 17 languages [18]. Until now, these associations have been modelled as pairwise relationships between words. More in detail, by construction, the recall of free associates always takes place in relationship with the same underlying stimulus [15]. Considering only pairwise relationships between the cue and its responses led to networks explaining the most variance in several lexical tasks (for details see [15]). Adding also pairwise relationships between responses themselves was shown to deteriorate network performance in explaining variance within lexical tasks and also added noise in the form of weak memory recall patterns between related responses [15]. In order to overcome noise, other techniques of pairwise network filtering, like maximal planar graph

embeddings or minimum spanning trees, have been successfully applied to free association networks (see [13, 21]). However, more work is needed to evaluate and understand the appropriateness of different network filtering technique [10]. Returning to the cognitive interpretation of one cue producing some activation signal stimulating recall of all responses at the same time [15, 23], thus giving rise to a higher-order interaction, we hereby propose a novel theoretical framework for modelling free association data: Cognitive hypergraphs.

Hypergraphs are complex networks where sets of nodes engage in the same (hyper)link simultaneously [25–27]. Whereas pairwise complex networks consider only links between two nodes, hypergraphs can consider connections among 3, 4 or more entities. In this way, hypergraphs can naturally encode for interactions between nodes of order higher than 2. This is strongly appealing for modelling free association data, as it enables for cue and responses to be combined together at the hyperlink level. The mathematics of hypergraphs originates from graph theory and combinatorics, with seminal work over graph isomorphism completed almost 40 years ago [25]. Only recently the formalism was extended by physicists and computer scientists to model a plethora of real-world complex systems [28, 29]. Marinazzo and colleagues used hypergraphs of information-theoretic associations between items in psychometric scales to reduce the impact of redundant information on identifying clusters of co-occurring symptoms compared to pairwise networks [30]. De Arruda and colleagues showed that analogous social contagion models on hypergraphs and pairwise networks would exhibit crucially different dynamics, with hypergraphs supporting critical phase transitions closer to empirical estimates and not reproduced by pairwise network structures [31]. Veldt and colleagues defined an affinity score for estimating homophily in groups, showing that in a scenario with 2 labels and equally sized hyperedges majority homophily can not be reached by both groups for a combinatorial impossibility of hypergraphs [32]. Sarker and colleagues extend the previous affinity score for groups with more than 2 labels and for simplicial complexes [33]. These examples are part of a quick multidisciplinary growth of data science models based on hypergraphs, which, however, contains a gap: Even comprehensive reviews of the field [26, 28] currently lack cognitive case studies.

To the best of our knowledge, our cognitive hypergraph framework represents a first-of-its-kind approach to modelling human memory and the mental lexicon through higher-order interactions [13] where concepts are represented as feature-rich nodes, i.e. nodes are endowed with vectors of psycholinguistic features [34]. The framework introduced here thus contains two points of novelty: (i) it combines response-response and cue-response beyond pairwise links through the mathematical formalism of hyperlinks; (ii) it enriches nodes with psycholinguistic features as to explore any interplay between higher-order interactions and conceptual features.

Focusing on sets of freely associated targets and cues as hyperlinks and including feature-rich representations of concepts/nodes, we explore and quantify the predictive power of cognitive hypergraphs against pairwise networks and standard psycholinguistic norms (neglecting any network structure) in reproducing word-level features. To do so we first extracted the +12,000 cue words from Small World of Words (SWOW) [18]. Next we determined the overlap with the words in the Glasgow lexico-semantic norms [3]. The resulting network consisted of cue-response pairs from SWOW for 3586 nodes. Each node was characterized by 11 features (i.e. covariates in psycholinguistic terms) representing

linguistic and psycholinguistic dimensions, namely valence, arousal, dominance, semantic size, concreteness, gender association, age of acquisition, familiarity, frequency, polysemy and length (see the Methods for descriptions of each). Within an interpretable machine learning framework, we aim to use either network or psycholinguistic features (or a combination of both), to predict a target covariate/feature of nodes. Emphasis is then given to comparing pairwise network features against hypergraph features or unstructured psycholinguistic norms. Interpretability [35] stems from the development of trained artificial intelligence (AI) models where the influence of one feature on model performance can be quantified and interpreted directionally (e.g. a higher feature improves regression performance). In this work, we focused on word concreteness as the predicted variable [2], using all others as predictor variables. We put emphasis over concreteness since it represents a crucial latent feature of words (not measurable directly like frequency or length [36]) that is vastly studied in cognitive neuroscience [37] and has been shown to affect several aspects of semantic cognition from lexical processing to information retention and knowledge internalisation [2].
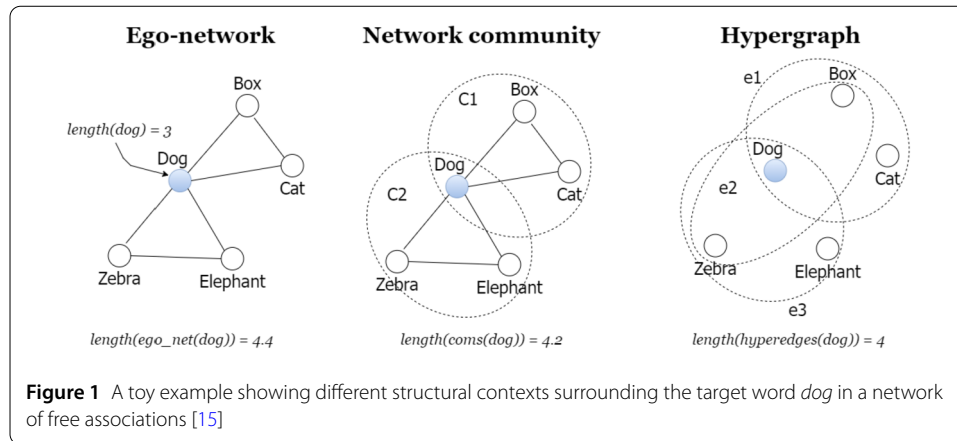
We provide new quantitative evidence that cognitive hypergraphs outperform both psycholinguistic baseline models and pairwise networks in predicting word concreteness from free association data. Our results underline the potential of going beyond pairwise interactions for modelling associative knowledge in human memory.

2 Results

We frame our analysis in the context of the studies about assortative mixing in the mental lexicon [16, 34, 38, 39]. Assortative mixing is an emerging behavior observed in many systems, such that nodes with similar features tend to connect together and stay apart from nodes with dissimilar features: The most common example refers to social networks, where individuals are more likely to interact in social circles if they share common features such as age, political, leaning, etc [40, 41]. Several studies propose a clustered mental lexicon such that groups of similarly concrete words would act as the building blocks of many cognitive processes, e.g., the formation of cue-response homogeneous patterns in memory recall [39]. Therefore, it would be possible to use the aggregated information provided by such groups to reconstruct/predict words' own traits, i.e., the empirical ground truth values according to a psycholinguist norm. For example, the concreteness of a word like "caterpillar" (i.e., its empirical ground truth value would be determined by words connected to it ("butterfly", "cabbage", etc). In the following, we discuss the rationale behind the adoption of several graph- and hypergraph-based representations for word associations (2.1), guided by psycholinguistic sources such as the Small World of Words (SWoW) project [18] and the Glasgow Norms [3] (2.2); finally, we discuss our main findings, namely that hypergraph-based modules of word associations overcome the other representations in the concreteness prediction task (2.3).

2.1 Rationale of aggregation strategies

Ego-networks   Figure 1 describes several word labeling procedures, i.e., the expression of a module/context by means of a characteristic value. We refer to a characteristic value of a context as the value associated to a target word as if that word was expressed by its direct (e.g. words directly linked) or indirect neighbors (e.g. words in the same community) rather than the word's own value. The example in the figure is based on the aggregation of one single feature, *length*, for one target word, *dog*. In Fig. 1 (left), we leverage

Figure 1 A toy example showing different structural contexts surrounding the target word *dog* in a network of free associations [15]

the ego-network of the free association network by just computing the average value of the feature, *length*, in the neighborhood of the target word, *dog*. In this way, the length of the target word will be 4.4 rather than 3 (as if the word was expressed by the ego-network context), being the former one the average of the word set context *box, cat, zebra, elephant* plus the target word itself, *dog*, included. The reason why we include the target word as well in the context-set is because the target word is an essential constituent of the semantic/conceptual context. Removing the target word from its own context would create a gap/hole in the structure itself that could model/imply undesirable or partial knowledge (cf. Appendix A), e.g. without the star centre an ego network would just be a collection of disconnected components. Importantly, the addition of the target word contributes only to the creation of an aggregate measure, influenced by indirect/direct neighbors and their properties (as contrasting with the properties of the target itself).

Contexts as local communities The aggregation based on the average value of nodes' ego-network is well-known and accepted in the literature of machine learning on graphs [42]. However, while reasoning about aggregation strategies in cognitive networks, one should consider that a word can be part of different contexts or neighborhoods [43, 44]. Hence, considering the whole ego-network could be an unsuitable proxy to estimate the value of a word *by the company it keeps* [43]. The free association network can still be used to identify more fine-grained contexts, e.g., the local communities surrounding a word [45, 46]. Figure 1 (center) shows a toy partition centered around the word *dog*. The free association graph structure unveils that the target word can participate in two different contexts/communities, $C1 = \{dog, box, cat\}$, and $C2 = \{dog, zebra, elephant\}$. This way the characteristic *length* value in *dog*'s context becomes the average of all the local communities/contexts where the word participates, 4.2.

Contexts as hyperedges However, contexts identified by ego-networks or network communities depend on an underlying network structure as a result of a heuristic process [18, 47]. Hence, we leverage the expressive power of hypergraphs to induce a higher-order context from the participant responses. Rather than creating several pairwise links between a cue and its responses, the hyperedges of a hypergraph can connect multiple elements simultaneously [26]. For each instance of the free association game, we model a hyperedge as the set that includes the cue word and all its responses. A response is thus modeled by means of a single connection rather than multiple pairwise links.

The characteristic value of a target word is calcuated as the average of the characteristic values of the hyperedges where the target word contributes in constituting an association pattern. In other words, while aggregating, we consider the so-called *star ego-network* of a target word in a hypergraph: from [48], the star ego-network of a node u in a hypergraph is defined as the set of all the hyperedges that include u. For the sake of simplicity, we do not consider here other connections among the connected hyperedges, as in other fine-grained definitions of higher-order ego-networks [48]. Let us discuss a brief example of the star ego-network.

Figure 1 (right) shows a set of responses involving the word *dog.* Three possible outcomes, i.e., hyperedges, indeed are $e1 = \{dog, box, cat\}$, $e2 = \{zebra, dog, box\}$, and $e3 = \{dog, zebra, elephant\}$. Word associations here are not constrained to pairwise relations only. For instance, in the toy association network there is no any direct link between *zebra* and *box.* This could happen for several reasons depending on the strategy used for reconstructing the graph. A possible explanation could be the following one. In the response *zebra, dog, box*, *zebra* is the cue word, *dog* is the first and *box* is the second response came to mind to the participant. Using a graph construction strategy where only consecutive words are connected, like a chain [18], *zebra* is not directly connected to *box*, but only indirectly connected through *dog.* Conversely, the hypergraph model merges all the three words by means of a single hyperedge. Doing so, the characteristic *length* value in *dog*'s context is not an average of all the graph-based contexts where the word participate (4.4, or 4.2) but an average of all its higher-order contexts, 4.

2.2 Setting the stage

Data overview We gain patterns for 3586 English words present both in the SWoW [18] and in the Glasgow norms [3] projects. From SWoW, we build the underlying graph/hypergraph structure; from the Glasgow Norms and other linguistic information easily available from words we form the vector of features to aggregate (cf. Sect. 4). Figure 2 provides a coarse-grained picture of the patterns emerging from different strategies. Each column provides an aggregation strategy. Each plot provides the characteristic values, except for

Figure 2 Scatter plots between the most important features according to the SHAP-values explanation (cf. Fig. 4). Each column represents an aggregation strategy, except for the first one. Points are always colored according to the original Glasgow Norms' concreteness
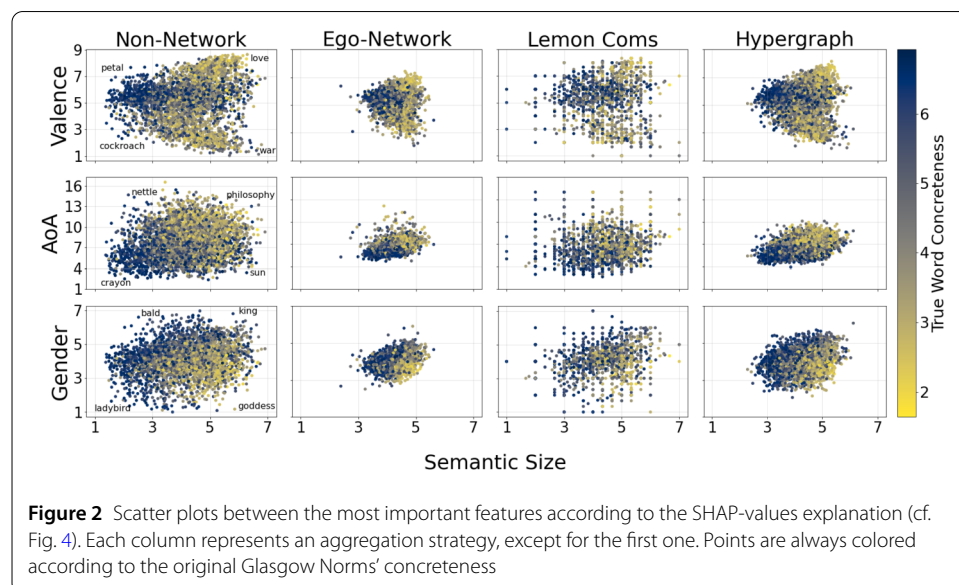
the first one, where each point describes the empirical ground truth value in the Glasgow Norms, e.g., *love* is an abstract (low concreteness) and salient (high semantic size) word associated to very positive emotions (high valence). In the second column, based on the ego-network strategy, the characteristic values result in a more flattened, overall compact cloud of points. Conversely, the hypergraph-based strategy comes as a hybrid between the non-network and the ego-network characteristics values, while the network community average values provide more coarse-grained value distributions (cf. later, Lemon communities).

Outline of aggregation algorithms　　Here is our methodology to extract/aggregate word features:

- *Non-Network*: No aggregation strategies are defined, i.e., we do not use any underlying structure from SWoW to extract a characteristic value;
- *Ego-Network*: Each word is described by a set of features whose characteristic value is the average of the word's ego-network (cf. 2.1);
- *Network communities*: We use different community-based strategies for feature aggregation; communities are found by using (i) a non-overlapping connectivity-based [49] community detection algorithm; (ii) a non-overlapping both connectivity- and feature homogeneity-based [16] algorithm; (iii) an overlapping local expansion method [46]; in detail:
  - i *Louvain* [49]: Same strategy as word's ego-network for aggregation. However, crisp communities provide larger contexts than ego-networks, since communities can group also nodes that are not directly neighbours [45]. The Louvain method is based on the family of algorithms that optimize the modularity function;
  - ii *Louvain "E"xtended to "V"ertex "A"ttributes (EVA)* [16]: Same strategy as word's ego-network for aggregation. EVA is an extension of Louvain that optimizes a linear combination of modularity and purity, a homogeneity-aware fitness function. Feature homogeneity-aware algorithms such as EVA force aggregations between words sharing similar feature values, in accordance with the word feature-homogeneity hypothesis [34, 39];
  - iii *Lemon* [46]: This strategy labels each target node with the average value of the local average context of the target word (cf. 2.1). The algorithm can capture small sets of overlapping communities. Rather than identifying a crisp/global structure, *Lemon* detects local modules given a representative set of seed nodes (cf. Materials and Methods). We run the algorithm *N* times, where in each run the seed node is a different word; this way, we can detect the local communities centered around all target words.
- *Hypergraph*: This strategy labels each target node with the average value of the hypergraph-based characteristic value contexts of the target word (cf. 2.1).

Details on prediction　　We test different algorithms from different families of methods to predict the concreteness value of a node.

- *Multiple Linear regression* [50]: Concreteness is expected to be a linear combination of the set of independent variables. The objective is to minimize the residual sum of squares between the observed targets (i.e., the original concreteness values) and the target predicted by the linear approximation;
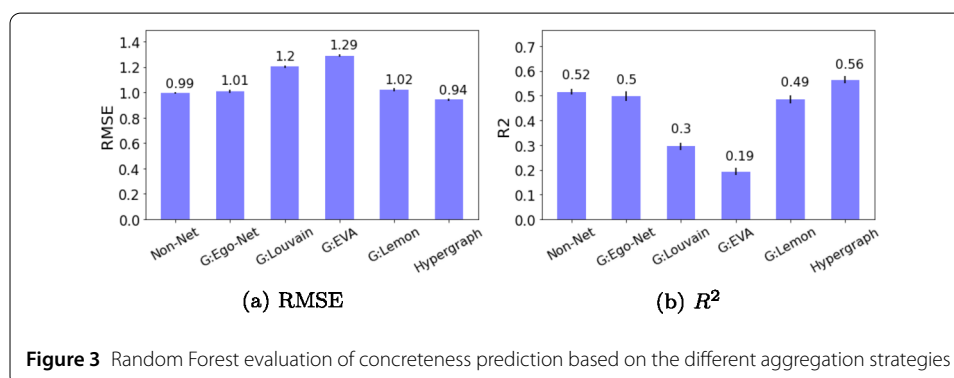
- *Random Forest* [51]: Several decision trees are built and the final output is based on the average of their predictions;
- *AdaBoost* [52, 53]: An ensemble method where a combination of weak estimators, e.g., decision stumps, are built sequentially to produce a stronger output;
- *Support Vector Machine* [54]: SVM's are used to find an appropriate hyperplane to fit the data while trying to define how much error is acceptable in the model.
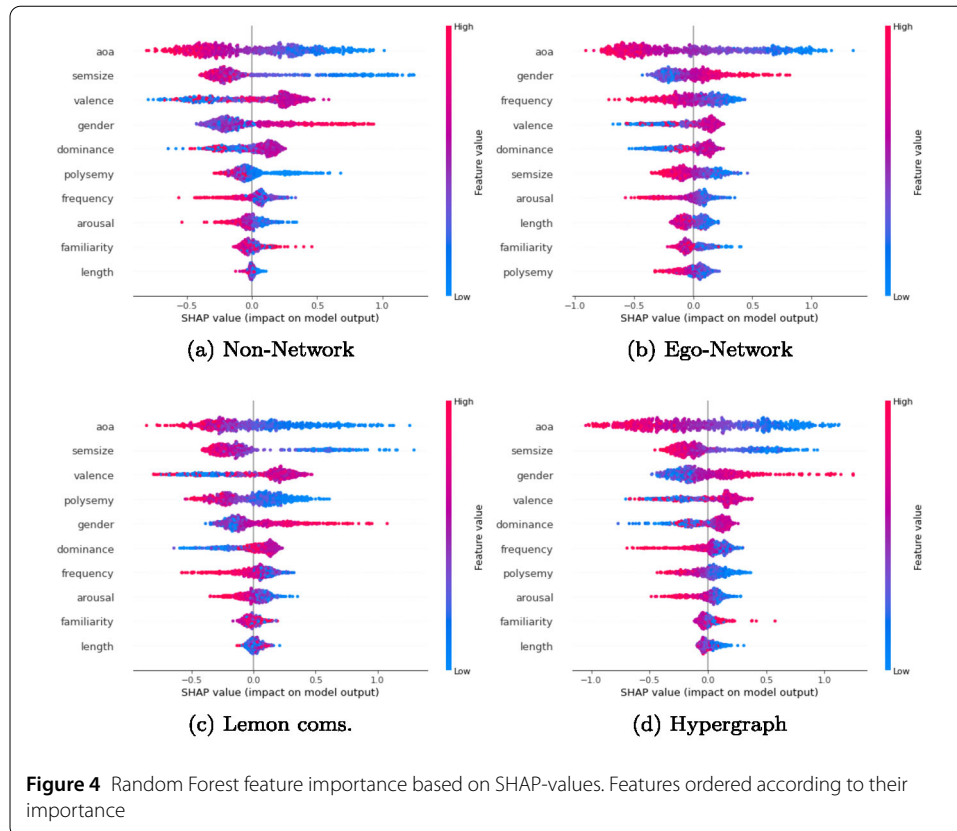
The algorithms provided similar results both in terms of evaluation performances and model explanation. We show in the main article the one that outperformed the others, the Random Forest (cf. Appendix B). Note that for each algorithm we provide hyperparameter tuning to maximize performances, and all the performance evaluations are cross-validated (cf. Sect. 4).

2.3 Predicting concreteness

We present here the Random Forest (henceforth, RF) performances on each dataset (cf. Sect. 4 for the RF hyperparameter tuning and Appendix B for other methods). The evaluation metrics in Fig. 3 highlight theperformances in terms of the average distance between predicted and original values, i.e., using the Root-Mean Squared Error (RMSE), and the variation in the variable in percentage terms (R^2). See *Materials and Methods, Evaluation details* for a precise description of the formulas. As can be seen from Fig. 3, the RF regressor provides better predictions on the set of features based on the hypergraph aggregation, while all the community-based strategies make the RF perform worse; performances on the ego-network aggregation and on the non-network strategy are similar. Different regression techniques are evaluated in Appendix B.

Figure 4 presents a more fine-grained evaluation based on feature importance with SHAP values [55, 56]. This evaluation highlights the impact of each feature on the estimation of word concreteness. A positive SHAP value for a datapoint (*x*-axis of each plot in Fig. 4) means that the predicted value on that datapoint is higher than a baseline predicted value – obtained in case that given feature was fixed to its expected value over the whole dataset —, and a negative SHAP value for a datapoint means that the predicted value on that datapoint is lower than the baseline. In other words, *x*-axis shows whether the effect of that feature value caused a higher or lower concreteness prediction. Thus, Fig. 4 shows that the RF predicts higher concreteness scores (on almost all the sets of features, net of different performances) when values of age of acquisition and semantic size are low, and when values of valence are high, as well as when words are associated with a masculine aspect of salience (high values of the gender variable). Conversely, the RF predicts lower

Figure 3 Random Forest evaluation of concreteness prediction based on the different aggregation strategies

Figure 4 Random Forest feature importance based on SHAP-values. Features ordered according to their importance
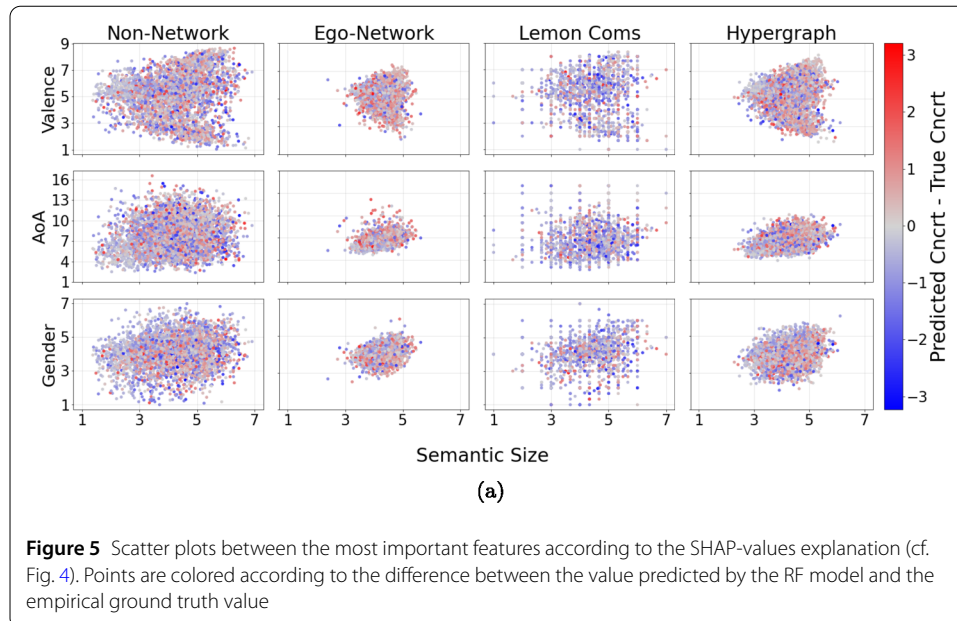
concreteness scores when values of age of acquisition and semantic size are high, when values of valence are low, and when words are associated with a feminine aspect of salience (low values of the gender variable).

To better understand these "profiles", let us focus again on Fig. 2. The scatter plots tell us there is correlation between concreteness and some other variables like valence, age of acquisition, gender and semantic size. For instance, there is a consistent group of early acquired, masculine-associated, concrete words with low values of semantic size and high values of valence. Also, there are some abstract words, i.e., words with low concreteness values, which are associated with medium-high values of semantic size. In fact, semantic size can be thought of as a proxy for conceptual salience across both abstract and concrete words, thus correlation with both concrete and abstract words is expected [57]. See *love* and *war*, for instance, which are two extremely high semantic salient words with opposite valence, where *love* is highly abstract, and *war* is highly concrete; cf. also *philosophy/sun* and *king/goddess* (cf. Fig. 2).

According to Fig. 2, the correlation remains unchanged in all the aggregation strategies. The combined results from Fig. 4 and Fig. 5 highlight that the RF can well predict a set of high concrete words associated with some characteristics such as early word acquisition or positive emotion. Figure 5 complements feature importance (cf. Fig. 4) and scatter plots (cf. Fig. 2) by coloring each word with respect to the residuals, i.e., the differences in the predicted and original concreteness. Note the "grey" zones, that indicate the words for which such differences in the predicted and empirical ground truth values are small: in this way, we can verify that the RF predicts the values of concrete words with the previous mentioned characteristics, validating the impact given by the SHAP values to profiles as

Figure 5 Scatter plots between the most important features according to the SHAP-values explanation (cf. Fig. 4). Points are colored according to the difference between the value predicted by the RF model and the empirical ground truth value

positive valence and early word acquisition (cf. Fig. 4). From Fig. 5, we can see that also abstract words can be well predicted by the RF; however, no clear patterns as the ones highlighted by SHAP summary plots emerge for the prediction of abstract words. Finally, Fig. 5 shows no noticeable variations in residuals across the different strategies. This indicates that the enhancement achieved through the utilization of hypergraph-based aggregation is attributable to improved regression (cf. Fig. 3) rather than the ability to predict specific profiles that cannot be captured by alternative aggregation methods or empirical ground truth values.

3 Discussion

Our work moves a step forward towards using hypergraphs [28] in cognitive modelling: Using hypergraphs provides richer cognitive measures compared to techniques that rely on communities or local neighborhoods. In other words, we show that the hypergraph formalism is better than pairwise networks or unstructured sets of features at predicting concreteness norms for individual words. Regression models on unstructured features try to predict a psycholinguistic norm of a target word/concept based on the word's own values, neglecting any conceptual association the target might have with other concepts. Why would connectivity matter? Recent work in cognitive network science has highlighted how memory recall patterns like the ones captured here can be highly insightful about semantic relatedness [21, 58], indicating that words separated by fewer memory recalls (i.e. shortest path length in terms of free associations) tend also to be rated as more semantically related. Shorter distance on free association networks thus corresponds to higher semantic relatedness.

Our working hypothesis is that the proximity between nodes in a semantic network translates into analogous values for mostly semantic psycholinguistic features, like concreteness [37]. Under this hypothesis, words closer to a target share similar concreteness norms and could thus enable quantitative predictions for the concreteness of the target itself. Consequently, our working hypothesis corresponds to the presence of a compart-

mentalisation of semantic features and network structure in the mental lexicon, where clusters of closer words can tend to share similar concreteness norms. Importantly, our work cannot identify a causal relationship, e.g. are the words connected because they are equally concrete, or are they concrete, because they have a certain number of connections? Despite this limit, our assumption identifies an insightful correlation. Network structure might thus be valuable for predicting the concreteness of one word by considering its close words/neighbours on a network topology of memory recall patterns. This hypothesis is supported by preliminary evidence in a previous work with pairwise network [23]. We test three ways for selecting neighbours to a given target word: (i) words linked to the target (i.e. network neighbourhood) based pairwise edges between cues and responses, (ii) words in the same community of the target in based on pairwise cue-response edges, and (iii) words linked to the target by sharing a hyperlink in a hypergraph representation of cue-response pairs. Notice that community analysis within the hypergraph representation of free associations [28] found trivial communities, which were discarded from the comparison.

We test our hypothesis through a machine learning framework. Model performance reports quantitative evidence that hyperlinks constitute the best proxy for predicting words' concreteness, outmatching both unstructured and structured models based on pairwise network neighbourhoods and communities.

These results confirm our working hypothesis and quantitatively indicate the presence of compartmentalisation in the layout of word associations that emerges more prominently when hypergraphs, rather than pairwise links in association graphs are considered. This clustering might emerge more in hypergraphs because they do not impose any specific distinction between the cue (e.g. "letter") and the responses (e.g. "mail", "sign", "dear"), which get represented within the same mathematical element (e.g. the hyperlink "letter", "mail", "sign", "dear"). In pairwise networks, instead, the cue is automatically a more relevant node than its responses [15], since the associations are encoded as links where the cue appears 3 times more frequently than the responses themselves, e.g. ("letter", "mail"), ("letter", "sign"), ("letter", "dear"). Not all words in free association networks are used as cues with the same frequency [18], this dichotomy leads to structurally different networks, whose predictive power of concreteness norms is different.
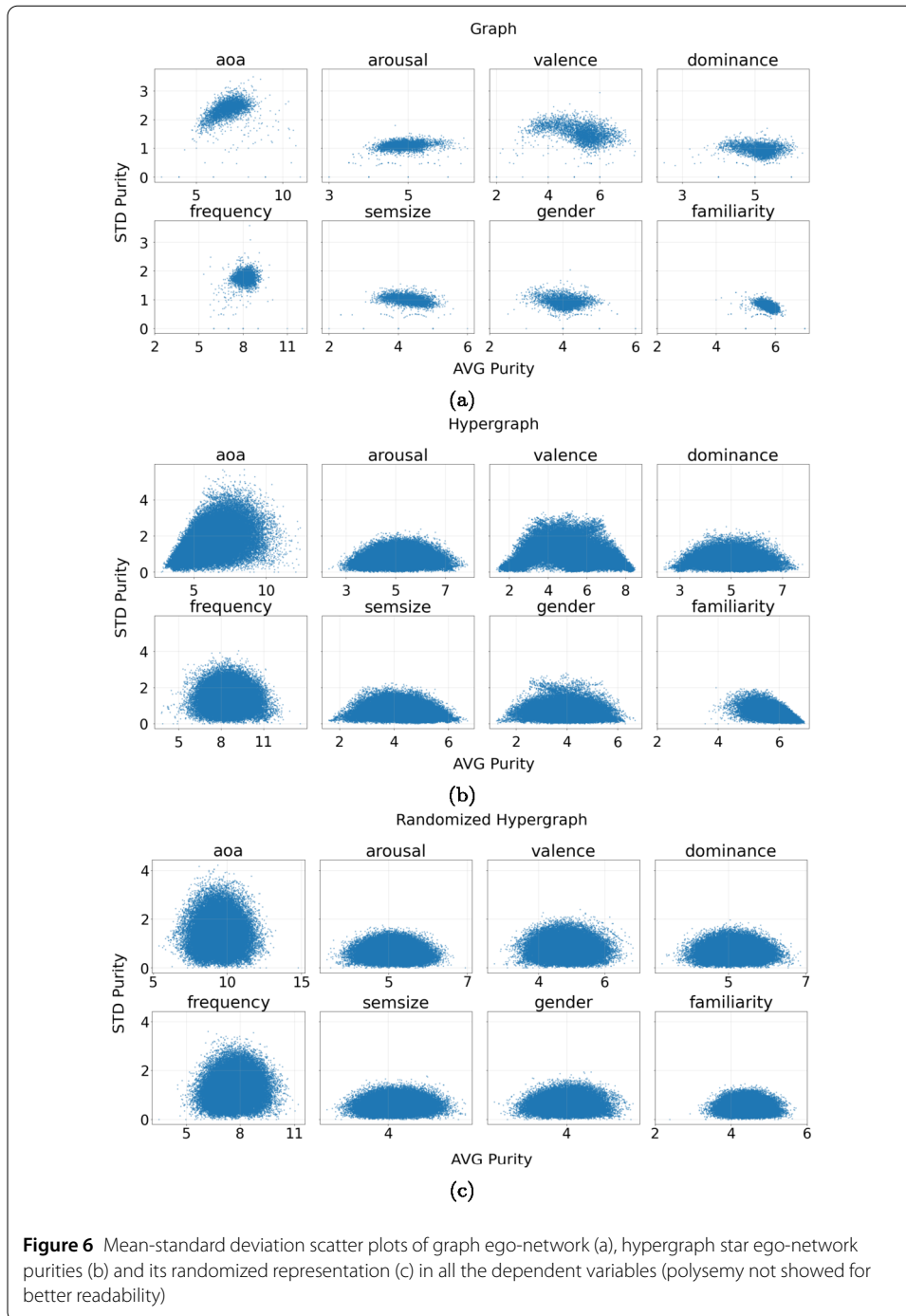
Cognitive hypergraphs represent a relatively novel tool for cognitive modelling because they are able to highlight a compartmentalisation phenomena that would be otherwise invisible with mainstream pairwise networks modelling free association data. Notice that we use the term "compartmentalisation" in a different way compared to previous approaches. In psychology, compartmentalisation is a strategy for separating conflicting and non-conflicting ideas [59]. We rather use this notion to identify a tendency for associative knowledge in the mental lexicon to form networked clusters/compartments of words sharing similar concreteness rates and appearing as being hyperlinked together. Unlike taxonomic categories, which are made of words sharing a common theme (e.g. all words being "animals" [60]), compartments identify coherence in terms of a semantic psycholinguistic feature (e.g. all words being highly concrete).

Our finding of feature-, hypergraph-based compartments in the mental lexicon agrees with previous works indicating a cognitive advantage in processing together more similar concepts [58, 61, 62]. Compartments might reflect a tendency for associative knowledge to be sorted in "patches" of concepts being thematically non-coherent but still similar in

terms of some psycholinguistic norms. In other words, compartments might reflect patterns of semantic foraging in the organisation and search of mental knowledge. Future research might investigate pre-existing frameworks for semantic foraging [61, 62] with novel contributions from hypergraphs. A challenge for this kind of research would be the assertion of which psycholinguistic features are mere consequences of more basic elements (e.g. frequency, length) and which are, instead, encoded properties of concepts, like concreteness, that cannot be fully explained by such basic elements only [63].

Notice that non-semantic psycholinguistic features might not give rise to compartmentalisation. In our tests, predicting a not purely semantic norm like the age of acquisition (AoA) of words (which does not depend only on semantics but also on phonological and orthographic features of words [64, 65]) resulted in regression models of unstructured norms behaving way better ($R^2 = 0.6 \pm 0.02$) than network-based pairwise (R^2 0.25 ± 0.02) and hypergraph ($R^2 = 0.45 \pm 0.03$) models (cf. Appendix C). Furthermore, hypergraph models behaved worse than unstructured norms even when predicting arousal, dominance, familiarity and length. Nonetheless, hypergraph models behaved significantly better (at least 5 times better in terms R^2) than pairwise network model in predicting these other 5 psycholinguistic dimensions. These differences are expected, since our working hypothesis relies on the finding that network distance reflects mostly semantic similarity. Non-semantic aspects of words might be affected in other ways by network structure, thus decreasing the performance of network-based models in predicting non-purely semantic norms (like AoA). When considering pairwise network, we can offer an intuitive argument about this lack of predictive power rising from network patterns. Previous works have shown that in pairwise networks non-semantic features follow disassortative rather than assortative patterns. Affective patterns like valence were shown to make pairwise free association networks become disassortative [20, 39], i.e. pairwise links connected words with opposite sentiment/valence polarities which often occur as antonym pairs (*pretty – ugly*, *young – old*) in free association pairwise networks. Disassortativity made pairwise network models powerful predictors of words' sentiment/valence [23], a pattern that we here explored under the framework of cognitive hypergraphs as introduced here. Cognitive hypergraphs surpassed both unstructured norms and pairwise networks in predicting valence (cf. Appendix C). This finding indicates that although parwise disassortative patterns exist in the network encoding of memory recalls, there is a stronger tendency for valence coherence to persist in subsequent recalls. Similarly to the mechanism of compartmentalisation we outlined above, this valence coherence creates clusters of words with similar valence and it cannot be captured unless one considers higher-order interactions, going from pairwise to hypergraph formalisms. Our findings thus indicate that non-semantic compartmentalisations can be noticeable in psycholinguistic data and push for more data-informed explorations of the organisation of psycholinguistic features within networks of memory recall patterns.

Compartmentalisation is present not only across the hyperlinks in a given neighbourhood but also among words within a single hyperlink. This tendency is even more evident for extreme values of norms. For instance, in Fig. 6(b), many hyperlinks tend to have words with similarly low age of acquisition norms. The extremes in Fig. 6(b) are not a statistical artefact when they cannot be reproduced by randomly sorting words in hyperlinks, which is the case for Fig. 6(c). This difference indicates a tendency for words in hyperlinks to be more similar in terms of age of acquisition, arousal, valence, dominance, semantic size,

Figure 6 Mean-standard deviation scatter plots of graph ego-network (a), hypergraph star ego-network purities (b) and its randomized representation (c) in all the dependent variables (polysemy not showed for better readability)

gender and familiarity when their average value for that norm is extreme, i.e. extremely low or high. This pattern further indicates a tendency for words to get compartmentalised even within hyperlinks and this might be due to an advantage in recalling concepts with similarly extreme psycholinguistic norms [61].

It has to be noted that compartmentalisation between concepts was quantitatively captured also by parallel distributed processing (PDP) models [66, 67]. PDP models quantify connections among individual features of each concept and then related knowledge retrieval to the strengths of the connections (e.g. the overlap in features) between elements

[68]. Despite this analogy, PDP models and cognitive hypergraphs adopt distinct representations of semantic memory. PDP models encode similarities in computational ways, so that concepts are related by means of a dynamical process or signal spreading across them [66]. Cognitive hypergraphs encode local relationships directly from empirical data, without needing additional computations. In this way, cognitive hypergraphs are more transparent than PDP models and can shed more light on the interplay between representational aspects of conceptual similarities and memory recalls, nonetheless PDP models can provide more insights about the dynamics of memory recall patterns and its failures [66, 69]. Future research could potentially merge representational and dynamical aspects of both modelling approaches to investigate memory recalls more closely.

In terms of limitations, one of the most important ones is relative to filtering free associations in hypergraphs. Firstly, Glasgow norms represent one among many repositories for psycholinguistic norms, see [2, 36, 65]. Based on the positive pioneering findings gained from this study, future research could test larger repositories of psycholinguistic variables that cannot be directly encoded in terms of network structure. The South Carolina Psycholinguistic Metabase (SCOPE) [70], which features 245 different lexical norms for 105,992 English words, represents a powerful candidate for future investigations with feature-rich hypergraphs, like the ones outlined here, and pairwise networks, like the ones investigated in [34]. Several prior works on free associations in pairwise networks have used some sort of filtering of infrequent or redundant word associations [20, 21]. Cognitive hypergraphs might not account for a statistical filtering of hyperlinks in some instances. In this dataset, applying the same statistical filtering introduced in [71], dismantled the whole set of hyperlinks. With link filtering being relevant for identifying meaningful network relationships and noisy links [13, 19], more techniques should be tested and designed in cognitive modelling settings. Another limitation of our approach revolves around a black-box nature of machine learning models [72], which are not yet commonly used in psychology. Black-box models make it difficult for the experimenter to identify how data is internally represented within the model, e.g. feature X being higher promotes the prediction of outcome Y. We try to address this issue by using Shapley values [73], a game-theoretic set of estimators for feature importance and contribution to model predictions. Although providing additional model interpretability, of relevance for cognitive modelling, Shapley values cannot provide causal evidence (feature X causes a better prediction of outcome Y) but only weaker correlation patterns [35]. Despite this, Shapley values were crucial to identify compartmentalisation in our data and should thus be more commonly used in future investigations merging artificial intelligence and cognitive modelling. Last but not least, this first-of-its-own investigation of cognitive hypergraphs as psychological models is indeed limited by the modest amounts of behavioural effects being considered here, i.e. the modelling presented here explored only free association data whereas modelling the mental lexicon might encompass multiple layers of behavioural data [14, 34]. This limitation is mainly due to the fact we focused our working hypothesis in terms of compartmentalisation within memory recalls only, without considering other psychological effects (e.g. reaction times in lexical decision-making tasks). Future works might explore whether the compartmentalisation found here could explain some variance in reaction times due to the dimensions that we found being well-captured by cognitive hypergraphs, i.e. concreteness and valence.

4 Materials and methods

Free associations The Small World of Words (SWoW) project[1] [18] is a large-scale database that aims to build mental dictionaries/lexicons in different languages from a word association test where each participant is asked to respond with at most 3 words coming to mind given a cue word. In this study we use the English lexicon (SWOW-EN), although other datasets in Dutch and Spanish are also available and new languages will be added in the future.[2]

Features The Glasgow norms [3] provide a multidimensional set of psycholinguistic variables describing a word in terms of emotion conveyed (valence, dominance), salience (semantic size, arousal, gender association), exposure (age of acquisition, familiarity), and visualization (concreteness). We use all the features available from this dataset except for age of acquisition, replaced with the data from [74], which provide more fine-grained information than the two-years binning from the Glasgow norm variable. Moreover, to increase the number of word dimensions, we also add information about word length, frequency and polysemy degree. Frequency is obtained from the OpenSubtitle dataset [75], and polysemy values are proxied by the size of the WordNet synsets [76]. A pre-processing step is needed before using the frequency variable, namely a logarithmic transformation, due to the well-known heavy-tailed distribution of this variable in human language [77]. Notice that when used for predictions, different variables are scaled to reduce normalisation issues.

Aggregation details For the creation of the free association network we strictly follow the *R123* procedure described in [18], namely that a link is formed between all the three responses and the cue word. Note that the responses are not connected in their turn to each other. The resulting graph $G = (V_G, E_G)$, with the filtering due to the matching between the SWoW and the Glasgow Norms words, has $V_G = 3586$ and $E_G = 165{,}690$. See also Appendix D for other pairwise-based aggregation strategies and the resulting graphs. The algorithms used for identifying communities depend on some parameters. A standard and accepted value of the resolution limit parameter γ is used for the Louvain algorithm, $\gamma = 1$. Moreover, the EVA algorithm, an attribute-aware extension of Louvain, also depends on a parameter α, that tunes the importance of forcing homogeneity within communities (the higher, the more homogeneous communities are identified). We set $\alpha = 0.8$ to obtain a partition significantly different from the Louvain one. Lemon is an algorithm from the family of seed set expansion methods, that neglect the global structure for identifying local modules expanding from a set of seed nodes. Usually, the seeding strategies involve random walks aiming to optimize some fitness score for communities [78, 79]. In detail, Lemon constructs the local spectra based on the singular vector approximations drawn from short random walks [46]. We use the original parameter values used in the Lemon algorithm paper [46], except for a preference on the maximum community size, set to 4 to explicitly simulate the set size of the SWoW responses.

Finally, the hypergraph $H = (V_H, E_H)$ resulting from the intersection between the SWoW and the Glasgow norms vocabularies has $V_H = 3586$ and $E_H = 67{,}600$.

Prediction details   In the RF model, we have chosen the best set of parameter values for the number of estimators (number of trees in the forest), the maximum number of features considered for splitting a node, the maximum depth, the minimum number of points placed in a node before the node is split, and the minimum number of points allowed in a leaf node. To find parameter values, we performed a 10-fold cross-validation, thus we evaluated average values and standard errors of RMSE and R^2 (cf. later) on the test sets of such 10 different splits of the data each time. After finding the parameters, for the sake of simplicity, we analyzed SHAP summary plots on a single data split in 80% train and 20% test. The whole prediction framework was implemented by considering the models, the methods, and the evaluation measures present in scikit-learn[3] and the SHAP library.[4]

Evaluation details   We evaluate the models with the root-mean-square error (RMSE) and the coefficient of determination (R^2).

To introduce RMSE, we first define the sum of the square of errors, or residual sum of squares, *RSS*, as follows:

$$RSS = \sum_{i}^{N} (y_i - \hat{y}_i)^2, \tag{1}$$

where N is the number of words, y_i is the empirical concreteness score of a word in the Glasgow Norms, and $\hat{y}_i$ is the score predicted by a model for that word. To understand this in our context, let us consider a model that predicts, respectively, a concreteness score of 6.5 and another of 4.5 for the two words *brain* and *mind*, which have, respectively, empirical ground truth values of 6.4 and 2.5 in the Glasgow Norms. The *RSS* is of 4.01, indicating there is, to some extent, some amount of error between the predicted and the empirical values. To better read the errors, it is often used RMSE, namely the square root of the average of *RSS*. Formally:

$$RMSE = \sqrt{\frac{1}{N} * RSS}. \tag{2}$$

In our toy example, the average of *RSS* is 2.005, thus *RMSE* = 1.41, indicating there exists variance in the predicted scores with respect to the empirical ground truth values.

Similarly, to describe R^2, we first introduce the total sum of squares, *TSS*, as follows:

$$TSS = \sum_{i}^{N} (y_i - \bar{y})^2, \tag{3}$$

where $\bar{y}$ is the average of the empirical ground truth scores, thus *TSS* sums over the squared differences between the empirical ground truth values and their average. R^2 is thus defined as follows:

$$R^2 = 1 - \frac{RSS}{TSS}. \tag{4}$$

---

[3]https://scikit-learn.org/stable/.

[4]https://shap.readthedocs.io/en/latest/.

In the example with the two words above, $\bar{y}$ is 4.45, and *TSS* is 7.6, and $R^2 = 0.47$. A different model that would predict a different value of the word *mind*, e.g., 2.8, would decrease *RMSE* and increase $R2$ for lower residuals.

Appendix A: Gap in hyperedges

An important point to discuss is the question whether including or not including the target word within the context in the hyperedge as well as in the local community obtained with the Lemon algorithm. We test such a choice within our machine learning framework in predicting concreteness, showing in Table A1 a decrease in the Random Forest performances on the Lemon- and hypergraph-based sets of features, where the target words are removed from their own contexts, simulating some kind of knowledge gap [80] in the memory recall patterns.

Appendix B: Performances of other models

As highlighted in the main text, the Random Forest predictor on the several different sets of features demonstrated that the hypergraph model achieves better results than the other aggregating strategies. To ensure that the result does not depend on a specific instance of a particular regressor, in Table B2 we show the performances of other predictors on the same sets of features. We perform a linear regression, as well as a Support Vector Machine model, and an ensemble method similar to the Random Forest framework but based on boosting. All the machine learning algorithms provide similar results such that the features based on the hypergraph aggregation continues to provide better performances in terms of RMSE and R^2. The only difference is in the magnitude of the scores, such that the

Table A1 Random Forest evaluation of concreteness prediction based on the Lemon and Hypergraph aggregation strategies with gaps, i.e., without the target word within the contexts

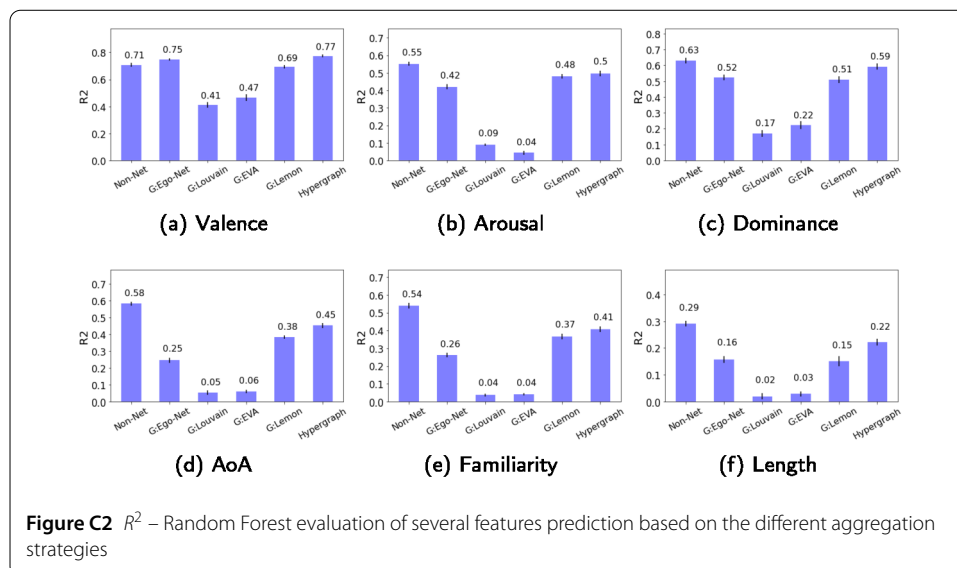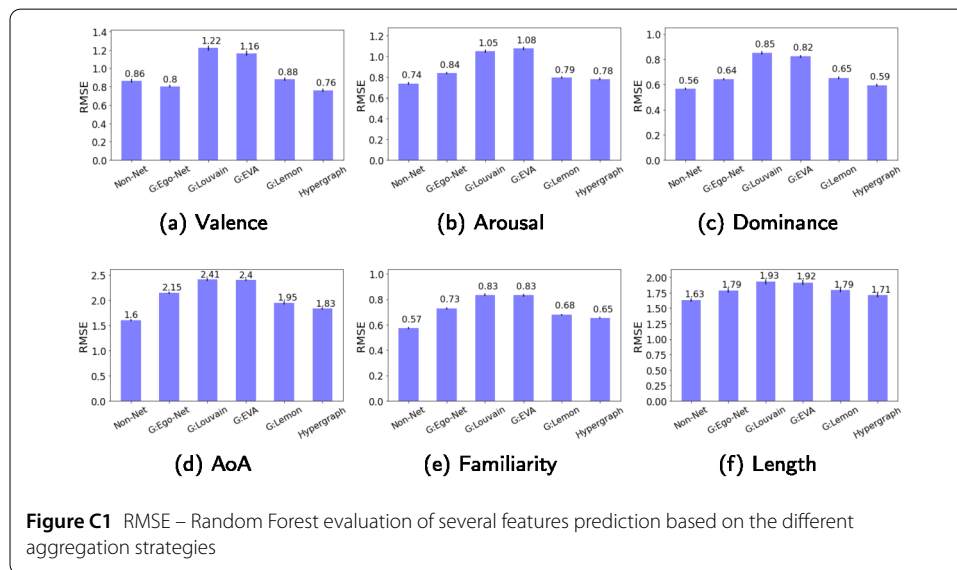|		G: Lemon (with gap)	Hypergraph (with gap)
RMSE	M	1.11	1.08
	SE	0.04	0.03
R^2	M	0.39	0.43
	SE	0.04	0.05

Table B2 Evaluation of concreteness prediction by different regression algorithms on the different sets of aggregation strategies

			Non-Net	G: Ego-Net	G: Louvain	G: EVA	G: Lemon	Hypergraph
Linear Regression	RMSE	M	1.17	1.09	1.45	1.45	1.18	1.08
		SE	0.04	0.04	0.04	0.04	0.04	0.04
	R^2	M	0.33	0.42	0.02	0.03	0.31	0.44
		SE	0.05	0.06	0.04	0.04	0.05	0.05
Support Vector Machine	RMSE	M	0.98	0.98	1.44	1.43	1.03	0.93
		SE	0.03	0.04	0.04	0.04	0.03	0.03
	R^2	M	0.53	0.53	0.06	0.05	0.48	0.58
		SE	0.04	0.06	0.03	0.03	0.05	0.04
AdaBoost	RMSE	M	1.13	1.10	1.45	1.45	1.16	1.06
		SE	0.03	0.02	0.04	0.05	0.03	0.03
	R^2	M	0.39	0.41	0.03	0.03	0.36	0.49
		SE	0.03	0.04	0.03	0.04	0.03	0.03

Random Forest performances, presented in the main article, are the highest among all the four regressors.

Appendix C: Predicting other features

As a main research subject for questioning network-based models of human memory, we limited our analysis in predicting concept concreteness. Figure C1 and Fig. C2 highlight a supplemental analysis, and show the results for the prediction of other features. Again, we compare the hypergraph strategy against the other graph-based and empirical representations already described in the main work. The same methodology for regression is applied as well, i.e., a hyperparameter-tuned Random Forest. We choose to compare the dimensions of valence, arousal, dominance, age of acquisition, familiarity and length, expecting different performances for them across the several aggregation strategies. Results tell us that, similarly to what we observed with concreteness, a hypergraph aggregation strategy

Figure C1 RMSE – Random Forest evaluation of several features prediction based on the different aggregation strategies

Figure C2 R^2 – Random Forest evaluation of several features prediction based on the different aggregation strategies

leads to better estimate valence, while the empirical values let the model perform better for all the other dimensions. As discussed in the main text, non-semantic psycholinguistic features might not give rise to compartmentalisation, as we particularly observe for AoA, familiarity, and length.

Appendix D: Other aggregation strategies

In this work, we tried to cover all the fundamental network-based aggregation strategies among pairwise ego-networks, graph communities and high-order ego-network representations, aiming to re-elaborate the features' values of a target word. However, other aggregation strategies may come to mind and, consequently, they may affect the results of a prediction. For instance, regarding the graph ego-network strategy, several other options are possible. In the main text, we represented the pairwise network using the so-called *R123* strategy, where links are placed between the cue word and the three responses, without connecting in their turn the responses (cf. *Materials and Methods, Aggregation details*). However, one might think that this strategy gives more importance to the cue word than to the responses. To validate the pairwise ego-network strategy, we also implemented other variants, particularly:

- the more straightforward *R1*, where the cue word is connected only to the first response;
- a variant where links are placed following a *chain*, e.g., the cue word is linked to the first response, then the second response is linked to the second response, etc;
- (iii) a variant where the cue word is linked to the three responses, and all the responses are in their turn connected to each other.

The last variant, in particular, can be thought of as another hypergraph-based strategy rather than a pairwise graph-based one, since each free association is represented as a clique. Also, we can distinguish the strategies according to the fact that some of them (R1 and R123) place edges between the cue word and the responses only, while other ones (chain- and clique-based) include edges between the responses as well, a procedure that gives more importance to the whole group.

The resulting graph $G_{R1} = (V_G, E_G)$, with the filtering due to the matching between the SWoW and the Glasgow Norms words, has $V_G = 3581$ and $E_G = 61{,}359$. Similarly, $G_{\text{Chain}} = (V_G, E_G)$ has $V_G = 3586$ and $E_G = 260{,}104$, and $G_{\text{Clique}} = (V_G, E_G)$ has $V_G = 3586$ and $E_G = 396{,}573$. Results are visible in Table D3. Note that values for R123 are the same presented in the main text. When only pairwise links between the cue word and the other responses are present (i.e., R1 and R123), results about concreteness prediction seem to be worse, while the performances improve when connections between responses are involved. These results suggest that, when connections between "implicit"/"indirect" words

Table D3 Random Forest evaluation of concreteness prediction based on alternative pairwise network constructions

		G: Ego-Net			
		No edges btw resp.		Edges btw. resp.	
		R1	R123	Chain	Clique
RMSE	M	1.03	1.01	0.95	0.94
	STD	0.03	0.02	0.03	0.05
R2	M	0.48	0.50	0.54	0.54
	STD	0.05	0.02	0.05	0.05

are placed, performances are better, a result that leads to consider the importance of compartmentalised models of free associations.

Abbreviations
AoA, Age of Acquisition; PDP, Parallel Distributed Processing; R^2, Coefficient of determination; RF, Random Forest; RMSE, Root-Mean-Square Error; RSS, Residual Sum of Squares; SCOPE, South Carolina Psycholinguistic Metabase; SHAP, SHapley Additive exPlanations; SWoW, Small World of Words; TSS, Total Sum of Squares.

Availability of data and materials
The original free associations analysed during the current study are available from the Small World of Words website: https://smallworldofwords.org/en/project/. The node covariates analysed during the current study are available from the Glasgow Norms paper [3]. Data preprocessing analysis is available at the following link: https://github.com/dsalvaz/Hypergraph-Cognitive-Networks.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author contributions
Conceptualization: SC, MS and GR; Data curation: SC, SD, MS and GR; Formal analysis: SC, MS and GR; Investigation: All authors; Methodology: SC, MS and GR; Supervision: MS and GR; Validation: SD and MS; Visualization: SC; Roles/Writing – original draft: All authors. All authors read and approved the final manuscript.

Author details
[1] Institute of Information Science and Technologies "A. Faedo" (ISTI), National Research Council (CNR), Via G. Moruzzi, 1, Pisa, Italy. [2] Computational Cognitive Science Lab, University of Melbourne, Melbourne, Australia. [3] Department of Psychology and Cognitive Science, University of Trento, Corso Bettini, 84, Trento, Italy.

References
1. Aitchison J (2012) Words in the mind: an introduction to the mental lexicon. Wiley, New York
2. Montefinese M (2019) Semantic representation of abstract and concrete words: a minireview of neural evidence. J Neurophysiol 121(5):1585–1587
3. Scott GG, Keitel A, Becirspahic M, Yao B, Sereno SC (2019) The Glasgow norms: ratings of 5500 words on nine scales. Behav Res Methods 51(3):1258–1270
4. Dóczi B (2019) An overview of conceptual models and theories of lexical representation in the mental lexicon. In: The Routledge handbook of vocabulary studies, pp 46–65
5. Vitevitch MS (2022) What can network science tell us about phonology and language processing? Top Cogn Sci 14(1):127–142
6. Castro N, Siew CS (2020) Contributions of modern network science to the cognitive sciences: revisiting research spirals of representation and process. Proc R Soc A 476(2238):20190825
7. Vitevitch MS, Ng JW, Hatley E, Castro N (2021) Phonological but not semantic influences on the speech-to-song illusion. Q J Exp Psychol 74(4):585–597
8. Wulff DU, De Deyne S, Aeschbach S, Mata R (2022) Using network science to understand the aging lexicon: linking individuals' experience, semantic networks, and cognitive performance. Top Cogn Sci 14(1):93–110
9. Valba O, Gorsky A (2022) K-clique percolation in free association networks and the possible mechanism behind the 7 ± 2 law. Sci Rep 12(1):1–9
10. Zock M, Ferret O, Schwab D (2010) Deliberate word access: an intuition, a roadmap and some preliminary empirical results. Int J Speech Technol 13(4):201–218
11. De Deyne S, Navarro DJ, Collell G, Perfors A (2021) Visual and affective multimodal models of word meaning in language and mind. Cogn Sci 45(1):12922
12. Kennington C (2021) Enriching language models with visually-grounded word vectors and the Lancaster sensorimotor norms. In: Proceedings of the 25th conference on computational natural language learning, pp 148–157
13. Siew CS, Wulff DU, Beckage NM, Kenett YN (2019) Cognitive network science: a review of research on cognition through the lens of network representations, processes, and dynamics. Complexity 2019:2108423
14. Stella M, Beckage NM, Brede M (2017) Multiplex lexical networks reveal patterns in early word acquisition in children. Sci Rep 7(1):1–10
15. De Deyne S, Navarro DJ, Storms G (2013) Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. Behav Res Methods 45(2):480–498

16. Citraro S, Rossetti G (2020) Identifying and exploiting homogeneous communities in labeled networks. Appl Netw Sci 5(1):1–20
17. Steyvers M, Tenenbaum JB (2005) The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. Cogn Sci 29(1):41–78
18. De Deyne S, Navarro DJ, Perfors A, Brysbaert M, Storms G (2019) The "small world of words" English word association norms for over 12,000 cue words. Behav Res Methods 51(3):987–1006
19. Kenett YN, Anaki D, Faust M (2014) Investigating the structure of semantic networks in low and high creative persons. Front Human Neurosci 8:407
20. Stella M, De Nigris S, Aloric A, Siew CS (2019) Forma mentis networks quantify crucial differences in stem perception between students and experts. PLoS ONE 14(10):0222870
21. Kenett YN, Levi E, Anaki D, Faust M (2017) The semantic distance task: quantifying semantic distance with semantic network path length. J Exp Psychol Learn Mem Cogn 43(9):1470
22. Stella M, Kenett YN (2019) Viability in multiplex lexical networks and machine learning characterizes human creativity. Big Data Cogn Comput 3(3):45
23. Vankrunkelsven H, Verheyen S, Storms G, De Deyne S (2018) Predicting lexical norms: a comparison between a word association model and text-based word co-occurrence models. J Cogn 1(1):45
24. Fatima A, Li Y, Hills TT, Stella M (2021) Dasentimental: detecting depression, anxiety, and stress in texts via emotional recall, cognitive networks, and machine learning. Big Data Cogn Comput 5(4):77
25. Berge C (1984) Hypergraphs: combinatorics of finite sets. In: Elsevier, vol 45
26. Battiston F, Amico E, Barrat A, Bianconi G, Ferraz de Arruda G, Franceschiello B, Iacopini I, Kéfi S, Latora V, Moreno Y et al (2021) The physics of higher-order interactions in complex systems. Nat Phys 17(10):1093–1098
27. Rosas FE, Mediano PA, Luppi AI, Varley TF, Lizier JT, Stramaglia S, Jensen HJ, Marinazzo D (2022) Disentangling high-order mechanisms and high-order behaviours in complex systems. Nat Phys 18(5):476–477
28. Battiston F, Petri G (2022) Higher-order systems. Springer, Berlin
29. Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young J-G, Petri G (2020) Networks beyond pairwise interactions: structure and dynamics. Phys Rep 874:1–92
30. Marinazzo D, Van Roozendaal J, Rosas FE, Stella M, Comolatti R, Colenbier N, Stramaglia S, Rosseel Y (2022) An information-theoretic approach to hypergraph psychometrics. arXiv preprint. arXiv:2205.01035
31. de Arruda GF, Petri G, Moreno Y (2020) Social contagion models on hypergraphs. Phys Rev Res 2(2):023032
32. Veldt N, Benson AR, Kleinberg J (2023) Combinatorial characterizations and impossibilities for higher-order homophily. Sci Adv 9(1):3200
33. Sarker A, Northrup N, Jadbabaie A (2023) Generalizing homophily to simplicial complexes. In: Complex networks and their applications XI: proceedings of the eleventh international conference on complex networks and their applications: COMPLEX NETWORKS 2022—volume 2. Springer, Berlin, pp 311–323
34. Citraro S, Vitevitch MS, Stella M, Rossetti G (2023) Feature-rich multiplex lexical networks reveal mental strategies of early language learning. Sci Rep 13(1)
35. Kumar IE, Venkatasubramanian S, Scheidegger C, Friedler S (2020) Problems with Shapley-value-based explanations as feature importance measures. In: International conference on machine learning, pp 5491–5500. PMLR
36. Brysbaert M, Warriner AB, Kuperman V (2014) Concreteness ratings for 40 thousand generally known English word lemmas. Behav Res Methods 46(3):904–911
37. Fliessbach K, Weis S, Klaver P, Elger CE, Weber B (2006) The effect of word concreteness on recognition memory. NeuroImage 32(3):1413–1421
38. Siew CS (2013) Community structure in the phonological network. Front Psychol 4:553
39. Van Rensbergen B, Storms G, De Deyne S (2015) Examining assortativity in the mental lexicon: evidence from word associations. Psychon Bull Rev 22(6):1717–1724
40. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27:415–444
41. Newman ME (2003) Mixing patterns in networks. Phys Rev E 67(2):026126
42. Bhagat S, Cormode G, Muthukrishnan S (2011) Node classification in social networks. In: Social network data analytics. Springer, Berlin, pp 115–148
43. Firth JR (1957) A synopsis of linguistic theory, 1930-1955. Studies in linguistic analysis
44. Lenci A (2018) Distributional models of word meaning. Annu Rev Linguist 4:151–171
45. Fortunato S, Hric D (2016) Community detection in networks: a user guide. Phys Rep 659:1–44
46. Li Y, He K, Bindel D, Hopcroft JE (2015) Uncovering the small community structure in large networks: a local spectral approach. In: Proceedings of the 24th International Conference on World Wide Web, pp 658–668
47. Zemla JC, Cao K, Mueller KD, Austerweil JL (2020) Snafu: the semantic network and fluency utility. Behav Res Methods 52(4):1681–1699
48. Comrie C, Kleinberg J (2021) Hypergraph ego-networks and their temporal evolution. In: 2021 IEEE International Conference on Data Mining (ICDM). IEEE, Los Alamitos, pp 91–100
49. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008(10):10008
50. Fisher RA (1922) The goodness of fit of regression formulae, and the distribution of regression coefficients. J R Stat Soc 85(4):597–612
51. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
52. Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55(1):119–139
53. Schapire RE (2013) Explaining adaboost. In: Empirical inference. Springer, Berlin, pp 37–52
54. Platt J et al (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Adv Large Margin Classif 10(3):61–74
55. Lundberg SM, Lee S-l (2017) A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds) Advances in neural information processing systems, vol 30, pp 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

56. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee S-I (2020) From local explanations to global understanding with explainable ai for trees. Nat Mach Intell 2(1):2522–5839

57. Yao B, Vasiljevic M, Weick M, Sereno ME, O'Donnell PJ, Sereno SC (2013) Semantic size of abstract concepts: it gets emotional when you can't see it. PLoS ONE 8(9):75000

58. Kumar AA, Balota DA, Steyvers M (2020) Distant connectivity and multiple-step priming in large-scale semantic networks. J Exp Psychol Learn Mem Cogn 46(12):2261

59. Ditzfeld CP, Showers CJ (2014) Self-structure and emotional experience. Cogn Emot 28(4):596–621

60. De Deyne S, Verheyen S (2015) Using network clustering to uncover the taxonomic and thematic structure of the mental lexicon. In: CEUR workshop proceedings, vol 1347, pp 172–176

61. Hills TT, Todd PM, Jones MN (2015) Foraging in semantic fields: how we search through memory. Top Cogn Sci 7(3):513–534

62. Todd PM, Hills TT (2020) Foraging in mind. Curr Dir Psychol Sci 29(3):309–315

63. Charbonnier J, Wartena C (2019) Predicting word concreteness and imagery. In: Proceedings of the 13th international conference on computational semantics-long papers. Association for Computational Linguistics, pp 176–187

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

65. Brysbaert M, Biemiller A (2017) Test-based age-of-acquisition norms for 44 thousand English word meanings. Behav Res Methods 49(4):1520–1523

66. Farah MJ, McClelland JL (1991) A computational model of semantic memory impairment: modality specificity and emergent category specificity. J Exp Psychol Gen 120(4):339

67. Rogers TT, McClelland JL et al (2004) Semantic cognition: a parallel distributed processing approach. MIT Press, Cambridge

68. Shabahang KD, Yim H, Dennis SJ (2022) Generalization at retrieval using associative networks with transient weight changes. Comput Brain Behav 5(1):124–155

69. Schapiro AC, Turk-Browne NB, Botvinick MM, Norman KA (2017) Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. Philos Trans R Soc B, Biol Sci 372(1711):20160049

70. Gao C, Shinkareva SV, Desai RH (2022) Scope: the south carolina psycholinguistic metabase. Behav Res Methods, 1–32

71. Musciotto F, Battiston F, Mantegna RN (2021) Detecting informative higher-order interactions in statistically validated hypergraphs. Commun Phys 4(1):1–9

72. Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1(5):206–215

73. Ghorbani A, Zou J (2019) Data Shapley: equitable valuation of data for machine learning. In: International conference on machine learning, pp 2242–2251. PMLR

74. Kuperman V, Stadthagen-Gonzalez H, Brysbaert M (2012) Age-of-acquisition ratings for 30,000 English words. Behav Res Methods 44(4):978–990

75. Barbaresi A (2014) Language-classified open subtitles (laclos): download, extraction, and quality assessment. PhD thesis, BBAW

76. Miller GA (1995) Wordnet: a lexical database for English. Commun ACM 38(11):39–41

77. Zipf GK (2016) Human behavior and the principle of least effort: an introduction to human ecology. In: Ravenio books

78. Whang JJ, Gleich DF, Dhillon IS (2013) Overlapping community detection using seed set expansion. In: Proceedings of the 22nd ACM international conference on information & knowledge management, pp 2099–2108

79. Kloumann IM, Kleinberg JM (2014) Community membership identification from small seed sets. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining, pp 1366–1375

80. Christianson NH, Sizemore Blevins A, Bassett DS (2020) Architecture and evolution of semantic networks in mathematics texts. Proc R Soc A 476(2239):20190741

Publisher's Note

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