Hypergraph models of the mental lexicon capture greater information for predicting the age of word acquisition

April 20, 2022

Abstract

Network science provides tools which can be used to model and study the structure of human memory and cognitive processes. In particular, the structure of language in the brain can be modelled using lexical networks. Most current approaches use simple pairwise connections between words, which ignores the complexity of language and the multiple simultaneous links between words and their meanings. In this work we overcome this limitation by modelling word associations from De Deyne et al.'s *Small World of Words* [1] as both a pairwise graph and a hypergraph of N=6003 words as nodes and, using centrality measures with machine learning algorithms, investigate whether there is any benefit to use of the hypergraph model over a traditional pairwise graph in predicting the age of acquisition of words. We additionally use this hypergraph model to predict the ordering with which N=497 words are acquired, using empirically observed age of acquisition data from English speaking toddlers, adding to work by Stella et al. [2]. We show that the hypergraph captures more information about the connections and so is more powerful than the pairwise graph for predicting age of acquisition norms in children. Our results open the way to novel approaches merging artificial intelligence and higher-order interactions for understanding cognitive development.

I certify that all material in this dissertation which is not my own work has been identified.

Contents

1 Introduction

Network science is the field concerned with understanding a complex system by analysis of the relations between its components. Complex systems appear seemingly everywhere that there are multiple connected parts. Our cities and towns are connected by vast networks of roads. The world wide web is an enormous complex network of its own [18], connecting websites and resources across the globe. Genes and proteins can be modelled as networks [19]. One can even imagine the network of all research papers, such as this one, with bibliographies describing the links between them. The study of network science is key to our understanding of a vast array of fields, and much can be drawn from analysing these different systems.

Network science is based upon the mathematical study of graph theory. A graph is made up of nodes (or vertices), which represent the individual units of the system, connected by edges, which represent the interactions between the units. The pattern with which these nodes are connected to each other represent the relationship between the units. In the context of the World Wide Web, for example, the nodes could represent websites, and the edges hyperlinks between them.

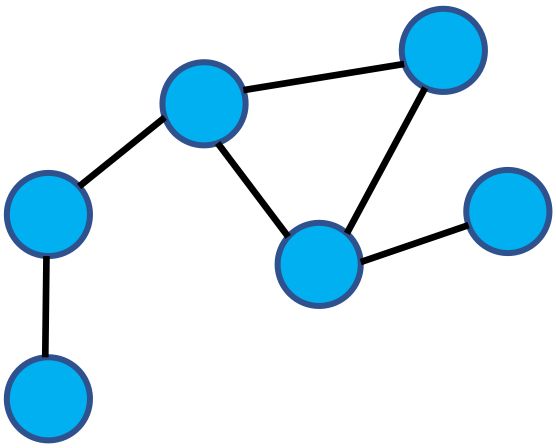

Figure 1.1: an example of a graph with six nodes.

An issue with traditional graphs using pairwise interactions is that it doesn't consider interactions that happen between more than two nodes simultaneously. For example, competition of resources in ecosystems happen between three or more species [8], social contagion in groups of people also function on a higher order [9]. These interactions can't be modelled by a traditional graph and so this means that graphs are potentially limited in the amount of information they are able to represent.

One way of representing these higher-order interactions is in the form of a 'hypergraph'. A hypergraph has nodes just like a traditional graph, but its 'hyperedges' can contain any number of nodes, instead of simply being connections between just two nodes [4]. This flexibility allows for more detailed and complex analysis of the interactions of parts of the system.

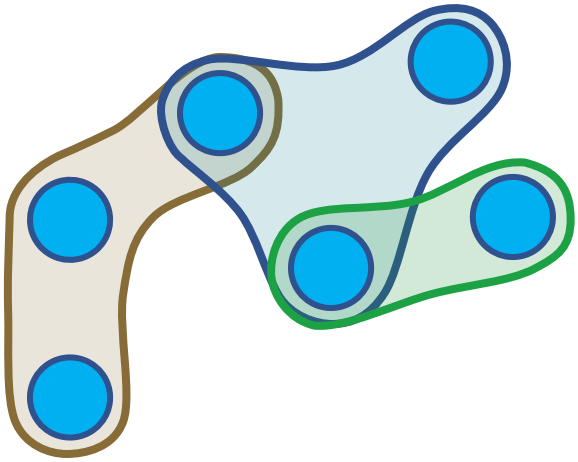

Figure 1.2: a hypergraph with six nodes and three hyper-edges.

Another example of a complex system is the *mental lexicon*. This is the network of interacting lexical items, such as words, that may be present in the brain. Much research has been done into trying to form an accurate image of the organisation of this network [20,23], but unfortunately one cannot directly observe this structure so there is disagreement amongst researchers about it. Understanding this structure is very important to understanding cognitive development in humans.

Most current research into modelling the structure of the mental lexicon focuses on using graphs with pairwise connections between the components [6,10,24]; this allows for only two items to interact with each other simultaneously and no more. However, due to the complex multi-relational nature of language, these approaches miss important information on the overall structure. Recently we have seen more studies into the use of higher-order network representations of semantic memory such as Stella et al.'s use of multiplex networks [2], and Liu and Cong's similar use of multi-layered networks [25]. This work will investigate a novel approach by using hypergraphs to represent semantic memory and, by analysis of the network, use machine learning algorithms to predict the age of acquisition of words.

2 Literature Review & Project Specification

2.1 Literature Review

The literature review looked at how research into the structure of semantic memory has evolved over the years, beginning with Quillian's [3] 1967 modelling of long-term memory as a network of words with conceptual connections between. He used this network to generate sentences, based on simple two-word inputs, that described the relationships between the two given words. These successful results suggested that his theory may be applicable to a real lexical network in humans. Then, in 1975, Collins and Loftus provided an attempt to apply Quillian's spreading activation theory of semantic memory to real results [4].

Later, in 2000, Brysbaert, De Deyne et. al studied the effects of the age of acquisition (AoA) of words on word processing tasks [5]. Using a word association task, they showed that AoA is an important factor of the speed for which word associations can be produced; words that are generally acquired early yielded faster responses for an associate word. Two experiments were conducted, and they concluded that the affect that AoA has on

word recognition is significant, however that more research was needed to determine if the effect is only on semantics or on more.

De Deyne and Storms [6] looked further in 2008 and investigated the properties of the results of a continuous word association task. They built and analysed a network of word associations and found a strong correlation between AoA and node centralities.

De Deyne et al. conducted a large collection of word association data and created, as of writing, the largest English-language resource for word associations norms in the world [1]. With over 12,000 cue words, the Small World of Words project provides a vast amount of data to be analysed. In their paper they revisited the idea of spreading-activation and showed that measures based on this mechanism were great predictors of direct judgements of similarity. This verifies some of Quillian's and Collin's and Loftus' ideas.

Hypergraphs also have their applications outside of cognitive network science, such as in Feng et al.'s use of hypergraphs in modelling gene expressions [7], successfully finding benefits of hypergraph models over standard graphs. Joslyn et al. used hypergraphs to represent, explore and analyse DNS data [17].

2.2 Project Specification

Two tasks will be completed to investigate the potential benefit of hypergraphs in the modelling of the mental lexicon. The first of these tasks is to train machine learning models to predict the age of acquisition of words using degree and closeness centrality data from the nodes on 1) a pairwise graph model and 2) a hypergraph model of the word association data described below. A random forest regressor will be used for this prediction task.

The second task is to train machine learning models to rank words in order of age of acquisition, also based on degree and closeness centrality measures from 1) a pairwise graph model and 2) a hypergraph model of the data described below. The graphs will be smaller than those of the first task as they will use a subset of the original word association dataset. An XGBoost ranker will be used here.

2.2.1 Dataset

The main data used in this project is a list of word associations from the *Small World of Words* (SWoW) project [1]. De Deyne et al. collected this data online from English speakers using a crowd-sourced approach. Data was provided by 88,722 volunteers with 62% identifying as female, 38% male, and <1% unspecified, and the average age of the participants was 36 years (with a standard deviation of 16). The participants were given the task online, wherein a cue word would appear on the screen, and they were required to enter the first three words that came to mind. Some biases must be noted, such as the fact that 62% of the respondents identified as female, as well as that 81% had at least a college or university bachelor's degree. Of the 13,000+ cue words, 6003 were used in the AoA prediction task.

Additionally, for the ranking task, an ordered list of 529 words was provided from Stella [2]. The list was obtained from The CHILDES Project dataset [21]. The words are ranked in the order of acquisition in English speaking toddlers, estimated by the parents. This was reduced to 497 words due to the other 34 words not being present in the SWoW data set.

2.2.2 Requirements

- A graph and a hypergraph representation of the free association data built in Python

- Graph and hypergraph models of a subset of the free associations, consisting of only the 497 words from Stella's list as mentioned above
- Calculations of degree and closeness centrality for each word in each graph model, using degree and s-closeness centrality for s=(1,2,3) on the hypergraphs
- Construction of random forest regressor models trained using the centrality measures for the AoA prediction task
- Construction of XGBoost ranking models trained using the centrality measures for the ranking task

Additionally, it will be interesting and beneficial to the research to investigate the use of other centrality measures such as betweenness centrality, and analyse the advantage this may give, though this will not be essential for the results.

2.2.3 Evaluation

To evaluate the model effectiveness, the following metrics will be used:

- Kendall's tau correlation coefficient
- Pearson's correlation coefficient (R)
- Root Mean Squared Error (RMSE)
- Coefficient of Determination (R^2)
- And for the ranking task: Normalised Discounted Cumulative Gain (NDCG)

These values will indicate the models' abilities at predicting the required data and will be important to make conclusions. Kendall's tau correlation coefficient measures the strength of association 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 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.

2.2.4 Tools

Python 3 was used along with various packages for building the graphs, implementing the machine learning models, and calculating various metrics. The first was NetworkX [15], a Python package that allows for the construction and analysis of graphs and networks. HyperNetX [16], a Python package for hypergraph analysis and visualisation, was also utilised. Finally, Scikit-learn [22] was used for the implementation of the machine learning models.

3 Graphs & Hypergraphs

We must first expand on the concepts of graphs and hypergraphs. A graph is a collection of nodes, or vertices, connected by edges. The nodes may have an unlimited number of connections between them, or none at all. Importantly, 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 figure 3.1.

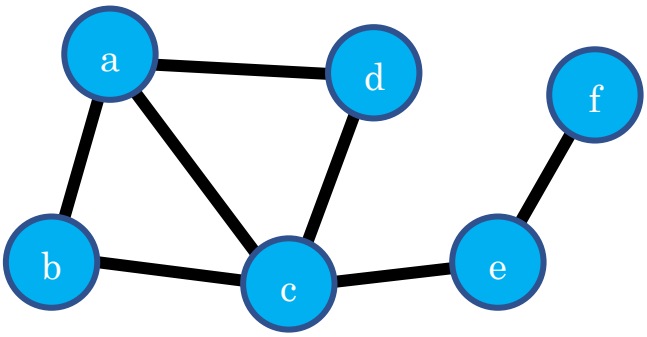

Figure 3.1: a graph with six nodes

We can write a graph with the notation $G = (V, E)$, where V is the set of nodes and E is the set of edges of the graph. The graph in figure 3.1 has node set $V = \{a, b, c, d, e, f\}$, and has edge set $E = \{[a, b], [a, c], [a, d], [b, c], [c, d], [c, e], [e, f]\}$.

To analyse the properties of the nodes in such a graph, we can look at various centrality measures which can show how 'central' a node is within the network. The two measures that will be used are the degree centrality, and the closeness centrality. A node's degree is simply measured by counting how many other nodes are connected to it. For example, in figure 3.1, the node labelled *a* has a degree of three as it is connected to the three nodes *d*, *b*, and *c*. Degree centrality gives a measure of how locally connected a certain node is within the graph.

To understand closeness centrality, we first define a *walk* on a graph as an ordered sequence of nodes, such that each pair of successive nodes are adjacent (i.e., they share an edge). Then the closeness centrality of a node i is calculated using the formula below [13,14]:

$$C(i) = \frac{N}{\sum_j d(i,j)}$$

Where N is the total number of nodes in the network, the function $d(i,j)$ is the shortest distance (or walk) between the two nodes i and j, and we sum over every node in the network. This gives a measure of how close a node is to every other node and is an important measure used in network science [10].

Hypergraphs expand on the traditional graph structure by allowing for higher-order interactions between more than two units by the use of hyperedges. A hyperedge can contain any number of nodes and a node can be contained in any number of hyperedges. In fact, the hypergraph is a generalisation of the graph, since a hyperedge with all edges being of size two would be equivalent to a graph.

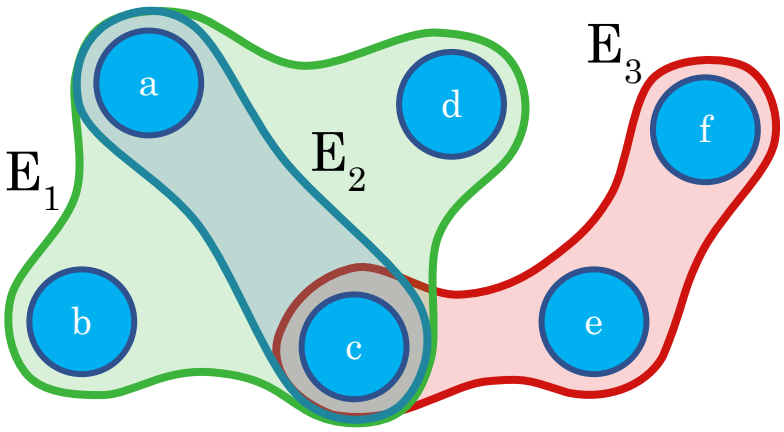

Figure 3.2: a hypergraph with six nodes {a,b,c,d,e,f} and three hyperedges {E_1,E_2,E_3}

For an example we look to figure 3.2 above. This hypergraph has six nodes connected by three hyperedges. It has node set $V = \{a, b, c, d, e, f\}$, and has edge set $E = \{E_1: [a, b, c, d], E_2: [a, c], E_3: [c, e, f]\}$. We can see that the nodes *a*, *b*, *c*, and *d* are all connected and contained within E_1, for example. Additionally, note that node *c* also appears in the hyperedges E_2 and E_3.

We can define degree and closeness centralities on hypergraphs also; we look to Aksoy's paper on hypernetwork science [11] for some definitions.

A node's degree in a hypergraph is simply a count of how many hyperedges contain it [11,13]. For example, in figure 3.2, the node *c* has a degree of three, as it appears within three distinct hyperedges; node *a* has a degree of two, and the other nodes all have degree one

Closeness centrality on a hypergraph is defined similarly to on a graph but with the addition of a parameter *s*. We define an *s-walk* as a generalisation of the walk on a graph that can apply to hypergraphs. Here the *s* parameter defines the minimum number of nodes in the overlap between two successive hyperedges [11]; for example, in a 2-walk there cannot be any movement between nodes where only a single node exists in the overlap, there must be at least two.

We will use the general definition of s-closeness:

$$C_s(i) = \frac{N}{\sum_{j \in E} d_s(i,j)}$$

Where $d_s(i,j)$ is the shortest s-walk between hyperedges *i* and *j,* and for N total nodes in the network. We sum over every hyper-edge in the edge set.

Using these measurements of degree and closeness on both graphs and hypergraphs provides the data needed to train the machine learning algorithms.

4 Methodology

The initial task was to use the word association data to predict the age of acquisition of words. To do so, a graph and a hypergraph representation of the same word association data will be made along with attempts at calculations of degree and closeness centralities of each node in the graph, and then these calculations will be used to train a machine learning model to make the predictions.

The second task is to train a ranker to predict the ordering with which words are learnt based on the 497 words from Stella [2]. Similar graph and hypergraph representations of the word association data in the first task will be made, however this time with a subset consisting only of the aforementioned 497 words. Degree and closeness centrality will also be used here to train the model.

4.1 Data Pre-processing

The raw word association data contained over one million rows, each row containing four separate 'word' entries. Unfortunately, not every entry contained an actual word as some participants appeared to have either misunderstood the task in some way or made a mistake with their entry. Examples of bad entries (sic) include:

- `h. p. lovecraft's fascination with non-euclidean geometry`
- `Barry at island pharmacy`
- `{sorry about that howdy do`

Additionally, there were many entries with redundant punctuation as well as phrases and typos which were not relevant to this task. These all had to be removed. At first the list was filtered using regex expressions to remove all these invalid/unhelpful entries. Eventually it was filtered to simply only include responses which contained one-word answers from the SWoW dataset.

4.2 Building the Graphs

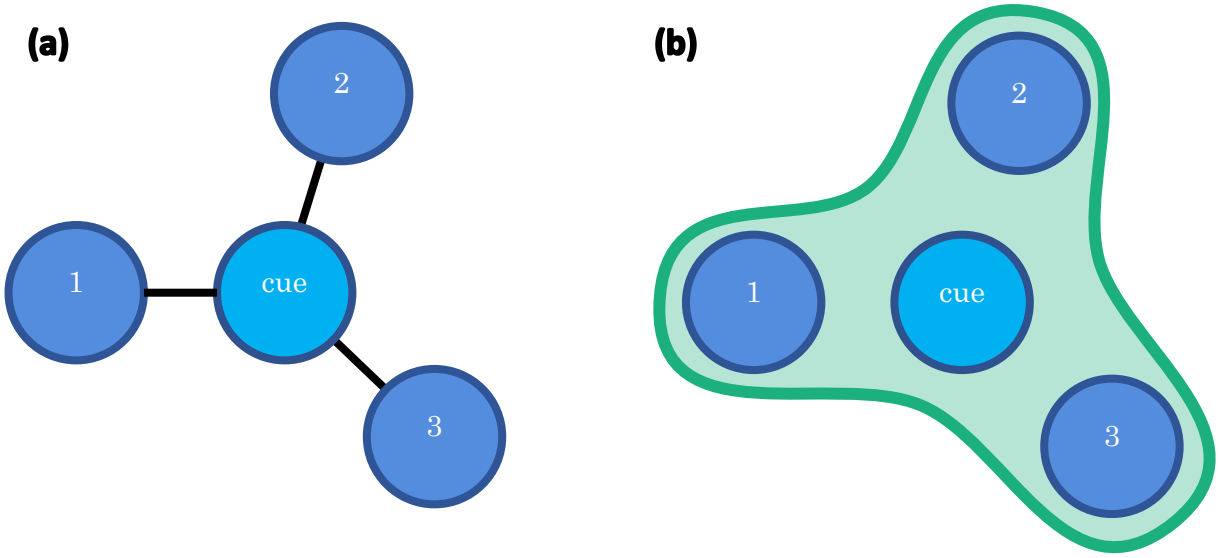

Figure 4.1. Left: a graph representation of the cue word and its three responses with individuals edges connecting cue and responses. Right: a hypergraph representation of the same, with a single hyperedge connecting the cue and all the responses simultaneously.

The NetworkX Python package was used to build the pairwise graph models of word associations, and HyperNetX for the hypergraph. The word associations were processed into a list of pairwise connections, where the cue word was connected to each of the subsequent user-inputted responses. These pairwise connections are then processed as either individual edges by NetworkX, or as hyperedges by HyperNetX. Both Python packages can then take the list of pairwise connections and build the relevant graph;

NetworkX builds us a graph with the cue and response words connected as in figure 4.1a, while HyperNetX constructs a hypergraph as in figure 4.1b.

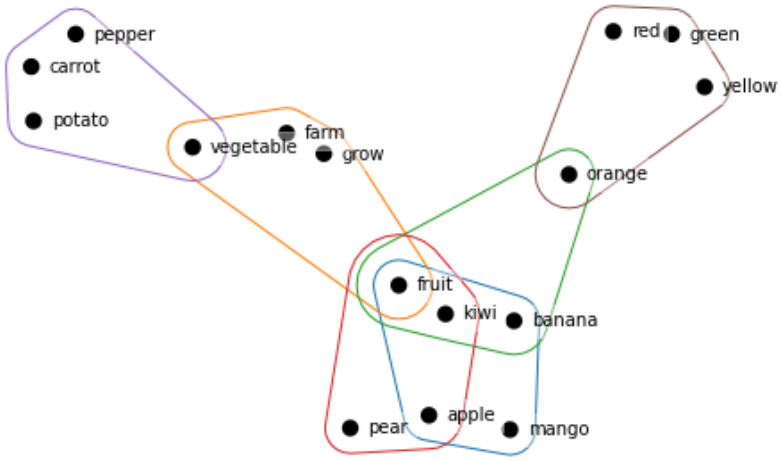

Figure 4.2: An example subgraph showing the hyperedges containing cue words and responses

An example of part of the hypergraph is shown in figure 4.2 for visual reference. The AoA prediction task used graphs with 6003 unique nodes, connected according to 207,915 4-word associations. For the ranking task we used a further subset of the word association data which included rows that contained only any of the 497 ranked words. This gave a list of 2649 4-word associations which resulted in a small enough hypergraph (497 nodes) so that the closeness centrality calculations were feasible.

4.3 Centrality Calculations

Initially the aim was to calculate degree and closeness centrality for each word in the graphs. In the AoA prediction task, the pairwise graph's degree and closeness centrality was calculated for each node, and for the hypergraph we would use degree and s-closeness centrality for s=1,2,3. However, the calculations proved to be very computationally demanding for the hypergraph due in part to the number of nodes (initially over 100,000) and possibly also due to HyperNetX's algorithms which may not be optimal for the calculations. We reduced the word association list to only include the rows that contained only words acquired before age 9, according to the AoA data. This reduced graph size allowed us to calculate the degree of each node in the hypergraph, so we decided to use just the degree for training the AoA prediction model.

In the ranking task, the hypergraph was small enough to be able to calculate the s-closeness centrality for each node which provided a vast amount more data to train the machine learning models with. The s-closeness (s=1,2,3) was calculated for each node using HyperNetX's built in function and stored alongside the degree calculations for training.

4.4 Machine Learning

The machine learning algorithm used for the AoA prediction was a random forest regressor, implemented with Scikit-learn. This was trained using the degree calculations, along with the word length and the logarithm of the frequency with which each word appeared in the word association data. The regressor was configured to use 150 trees and a mean squared error criterion for measuring the quality of a split.

For the ranking task we used an XGBoost ranker, also implemented with Scikit-learn. The ranker was configured to use 150 estimators and a learning rate of 0.2 and was trained using degree and (s-)closeness centrality. The logarithmic frequency and word length were not used here, in order to gain an understanding of how only the graph structure influences prediction quality.

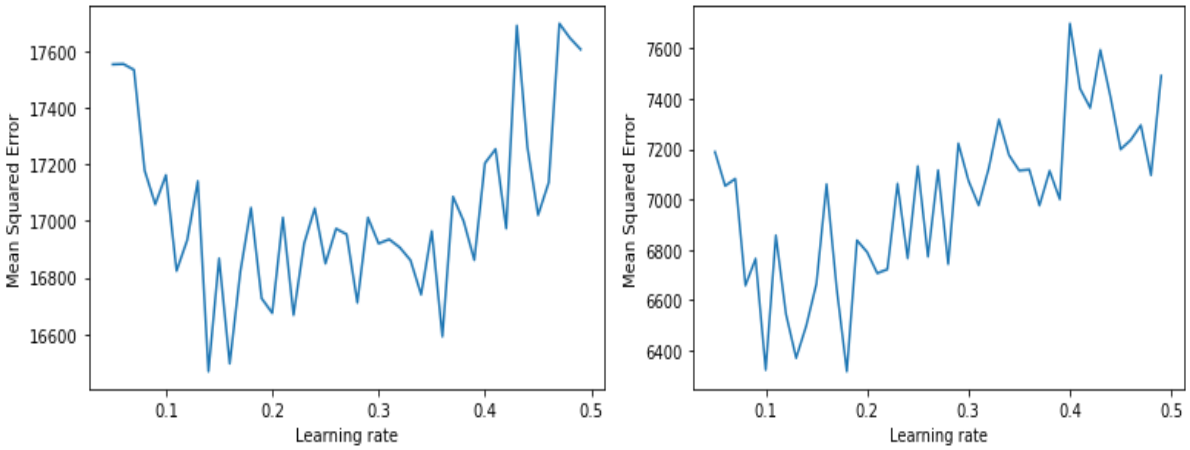

Figure 4.3: Learning rate vs. error plots used for parameter tuning of the ranking model. Left: results with the pairwise graph model. Right: results with the hypergraph model.

The learning rate was selected by computing the error rate with a varied learning rate and choosing a value that gives the smallest error. Figure 4.3 shows graphs of error against learning rate, showing that a value of 0.2 gives roughly the lowest error on both models.

4.5 Feature Importance Testing

When working with machine learning models it's important to analyse the effect that each feature has on the performance of the model. 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 metrics calculated. This process was repeated 100 times to gain average values of the change in metrics for each shuffling. A table of results can then be constructed for both graph and hypergraph models, as shown below.

4.5.1 Age of Acquisition Prediction

In the AoA prediction there were three features used: degree, word length, and the logarithm of word frequency. Table 4.1 shows the results for the pairwise graph and table 4.2 shows that for the hypergraph.

Metric	Original	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 4.1: Pairwise graph feature importance results. *The largest absolute difference for each metric is highlighted in bold.*

Metric	Original	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 4.2: Hypergraph feature importance results *with largest absolute differences highlighted in bold.*

Tables 4.1 and 4.2 show how the performance metrics of the two graph models changed with the shuffling of each feature. In both cases shuffling the degree column had the greatest impact, with a 6.3% increase in root mean squared error on the graph model and a 12.7% increase on the hypergraph model. Clearly the degree was the most important feature for this prediction task; this is not a surprising result considering that the degree was the only value derived directly from the graphs, providing information on the individual nodes and their structure. The shuffling of word length had a very minor effect. 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, with a p-value of 0.0315 for the hypergraph and a p-value of 0.000856 for the pairwise graph. Therefore, it appears that word length does have a small effect on the prediction quality.

4.5.2 Ranking

The ranking models used degree and closeness centrality to make predictions, with the hypergraph model using s-closeness for s=(1,2,3).

Metric	Original	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**

Table 4.3 Pairwise graph feature importance results with the largest difference in metrics bolded

Metric	Original	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

Table 4.4 Hypergraph feature importance results with the largest difference in metrics bolded

Table 4.3 clearly shows that the 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. This implies that closeness centrality was the strongest predictor of the word ordering.

The results for s-closeness centrality in table 4.4 were averaged over to reduce the table size. 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, which agrees with the table 4.3.

5 Results

In this section we show and discuss the results of the models. First the AoA prediction results and then the ranking task results.

5.1 Age of Acquisition Prediction

The AoA prediction was successful in showing the benefit of the hypergraph in modelling the mental lexicon. Neither model was very strongly predictive of AoA, however the hypergraph model consistently performed better than the graph model. Table 5.1 and

figure 5.1 below show the results for both models. These results are the averaged set of results from 100 individual trainings of the prediction models, each trained on random splits of training/testing data (80% training/20% testing).

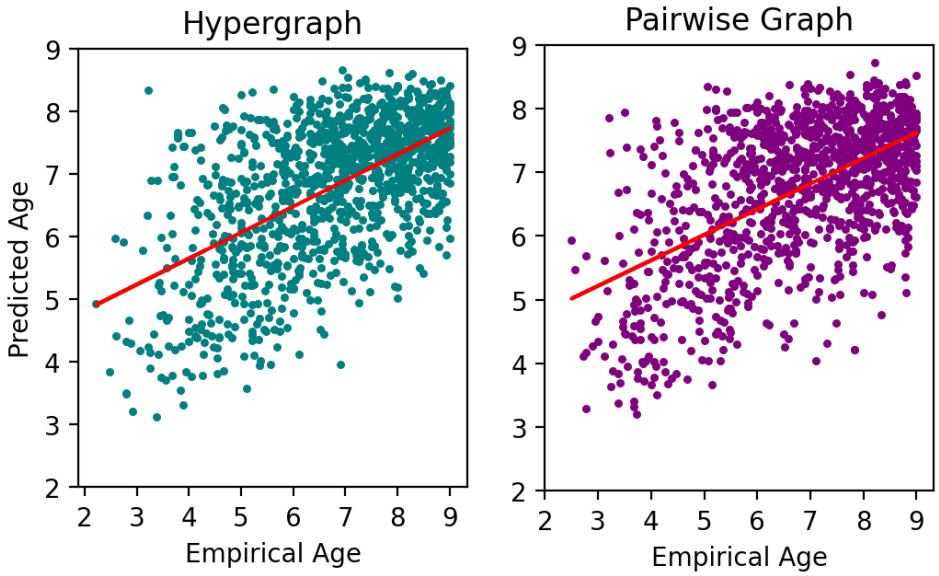

Figure 5.1: Graphs showing the relationship between predicted and empirical age for both hypergraph and graph models.

Metric	Hypergraph	Graph
Root Mean Squared Error	**1.2440**	1.3217
Kendall tau	**0.3832**	0.3181
Pearson's r	**0.6057**	0.5285
R^2	**0.3521**	0.2662

Table 5.1: AoA prediction random forest metrics.

The hypergraph model outperformed the graph model by a significant margin. It gave a 6% lower error rate and a Kendall tau correlation value 20% higher.

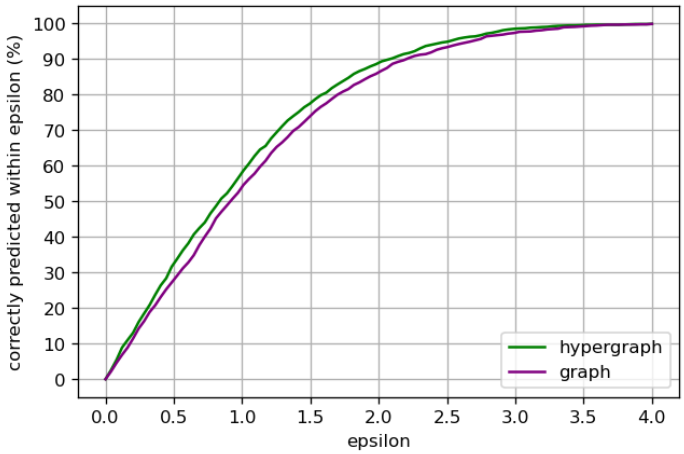

Figure 5.2: Percentage of results that were correctly predicted within ±epsilon of the empirical value.

Figure 5.2 above 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.

These results clearly show that the hypergraph model was more predictive of AoA than the pairwise graph model. As the degree was the only feature used for training directly derived from the graphs, we can conclude that the degree on the hypergraph provides information that better correlates with AoA than the degree measured on the pairwise graph.

5.2 Ranking Task

The ranking task was very successful at showing the benefit of the hypergraph model of the mental lexicon. A very strong correlation can be seen in figure 5.2 for the hypergraph model. These results are the averaged set of results from 100 individual trainings of the prediction models, each trained on random splits of training/testing data (80% training/20% testing).

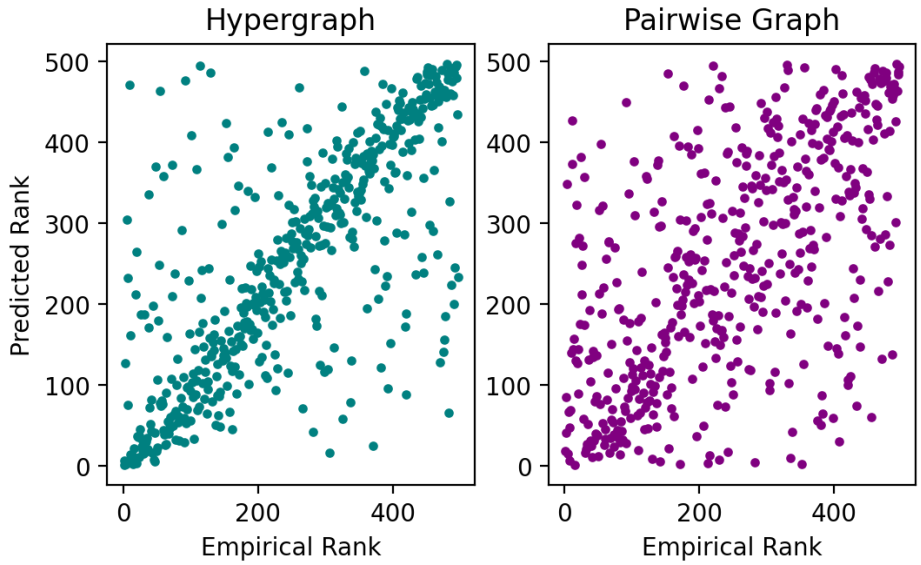

Figure 5.3: Graphs showing the relationship between predicted and true rank in both hypergraph and pairwise graph models.

Metric	Hypergraph	Graph
Root Mean Squared Error	**82.96**	129.86
Kendall tau	**0.6969**	0.4254
Pearson's r	**0.8328**	0.5904
R^2	**0.6656**	0.1807
Normalised Discounted Cumulative Gain	**0.9824**	0.9589

Table 5.2: Ranking prediction model metrics.

The hypergraph model significantly outperformed the pairwise graph model in the ranking task. It provided a visibly more correlated graph (figure 5.3), and this is reflected

in the results in table 5.2, 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.

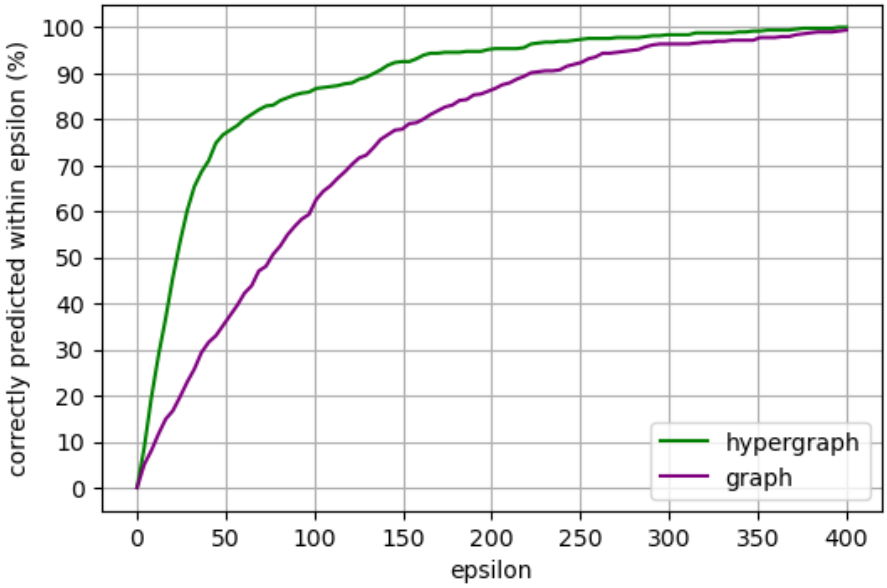

Figure 5.4: Percentage of results that were correctly predicted within ±epsilon of the empirical value.

The graph in figure 5.4 above shows the benefit of the hypergraph model, which made 77.1% of the predictions within ±50 positions of the empirical rank, whereas the graph model predicted only 36.6% in the same range.

These results show the strong benefit of the hypergraph model in prediction accuracy for this ranking task. There is no doubt that the hypergraph provided more information that correlated with the ordering of words, through the use of s-closeness centrality as well as degree.

6 Conclusions

Our results clearly show the benefit of the hypergraph model. Even through just the use of the 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 the 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 strong results, predicting over two times as many results in the range of 50 ranks of the correct value and providing an error rate 36% lower.

It seems that the hypergraph is able to capture more information on the structure of the mental lexicon in regard to age of acquisition and this research provides promise for the future use of hypergraphs in this setting.

Not all the initial goals were met in this project. Firstly, 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. The results gained were still conclusive enough, however would likely be even stronger given the use of more of the dataset. Additionally, for the same reason of computational power, the closeness centralities could not be calculated on the hypergraph for the first prediction task. Considering, however, that the results still showed the benefit of the hypergraph model this may be seen as a good thing as we can see objectively that the degree alone on the hypergraph was a more predictive feature than the degree on the pairwise graph; with closeness centrality involved, this might have been missed. Finally, the use of other centrality measures (such as betweenness centrality) were not investigated. This didn't hinder the results but could've provided further evidence for the benefit of the hypergraph.

Overall, the results are conclusive in showing that a hypergraph model of the mental lexicon can provide more intricate structural information that benefits a machine learning algorithm in its predictions of age of acquisition.

7 Future Research

There are some improvements which could be made to further the results of this project. The first being to use more of the *Small World of Words* dataset in building the graphs. Computational limitations restricted the use of more of the data, though we still gained conclusive results with the fraction used. With a more powerful machine larger graphs may be built. HyperNetX provides an add-on, NWHy, for use in Linux environments that supports optimised methods, including for s-centrality measures, which could be used to increase efficiency. Alternatively, the group that developed HyperNetX also provides a library for the high-performance parallel computing language Chapel, called CHGL (Chapel HyperGraph Library) [26]. With the use of supercomputers, the potential for vast amounts more data gathered from hypergraph structure may be possible.

8 References

[1] 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. *Behav Res* 51, 987–1006.

[2] 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.

[3] Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral science*, *12*(5), 410-430.

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

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

[6] De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior research methods*, *40*(1), 213-231.

[7] Feng, S., Heath, E., Jefferson, B., Joslyn, C., Kvinge, H., Mitchell, H. D., ... & Purvine, E. (2021). Hypergraph models of biological networks to identify genes critical to pathogenic viral response. *BMC bioinformatics*, *22*(1), 1-21.

[8] Levine, J. M., Bascompte, J., Adler, P. B., & Allesina, S. (2017). Beyond pairwise mechanisms of species coexistence in complex communities. Nature, 546(7656), 56-64.

[9] de Arruda, G. F., Petri, G., & Moreno, Y. (2020). Social contagion models on hypergraphs. *Physical Review Research*, 2(2), 023032.

[10] Goldstein, R., & Vitevitch, M. S. (2017). The influence of closeness centrality on lexical processing. *Frontiers in psychology*, *8*, 1683.

[11] 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.

[12] Lu, L., & Peng, X. (2011, May). High-ordered random walks and generalized Laplacians on hypergraphs. *International Workshop on Algorithms and Models for the Web-Graph* (pp. 14-25). Springer, Berlin, Heidelberg.

[13] 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

[14] Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, *1*(3), 215-239.

[15] Hagberg, A., Swart, P., & S Chult, D. (2008). *Exploring network structure, dynamics, and function using NetworkX*.

[16] HyperNetX. (2021). Retrieved from https://github.com/pnnl/HyperNetX. First accessed 16 Nov 2021.

[17] Joslyn, C. A., Aksoy, S., Arendt, D., Firoz, J., Jenkins, L., Praggastis, B., ... & Zalewski, M. (2020, September). Hypergraph analytics of domain name system relationships. In *International Workshop on Algorithms and Models for the Web-Graph* (pp. 1-15). Springer, Cham.

[18] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., ... & Wiener, J. (2011). Graph structure in the web. In *The Structure and Dynamics of Networks* (pp. 183-194). Princeton University Press

[19] Solé, R. V., & Pastor-Satorras, R. (2002). Complex networks in genomics and proteomics. *Handbook of Graphs and Networks*, 145-167.

[20] Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1-3), 1-22.

[21] MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.

[22] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

[23] Sanches, C., Routier, A., Colliot, O., & Teichmann, M. (2018). The structure of the mental lexicon: What primary progressive aphasias reveal. *Neuropsychologia*, *109*, 107-115.

[24] Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research, 51*(2), 408–422.

[25] Liu, H., & Cong, J. (2014). Empirical characterization of modern Chinese as a multi-level system from the complex network approach. *Journal of Chinese Linguistics*, *42*(1), 1-38.

[26] Jenkins, L., Bhuiyan, T., Harun, S., Lightsey, C., Mentgen, D., Aksoy, S., Stavcnger, T., Zalewski, M., Medal, H. & Joslyn, C. (2018, September). Chapel hypergraph library (chgl). In *2018 IEEE high performance extreme computing conference (HPEC)* (pp. 1-6). IEEE.