Understanding Higher-Order Correlations Among Semantic Components in Embeddings

Momose Oyama[1,2] **Hiroaki Yamagiwa**[1] **Hidetoshi Shimodaira**[1,2]
[1]Kyoto University [2]RIKEN
oyama.momose@sys.i.kyoto-u.ac.jp, hiroaki.yamagiwa@sys.i.kyoto-u.ac.jp,
shimo@i.kyoto-u.ac.jp

Abstract

Independent Component Analysis (ICA) offers interpretable semantic components of embeddings. While ICA theory assumes that embeddings can be linearly decomposed into independent components, real-world data often do not satisfy this assumption. Consequently, non-independencies remain between the estimated components, which ICA cannot eliminate. We quantified these non-independencies using higher-order correlations and demonstrated that when the higher-order correlation between two components is large, it indicates a strong semantic association between them, along with many words sharing common meanings with both components. The entire structure of non-independencies was visualized using a maximum spanning tree of semantic components. These findings provide deeper insights into embeddings through ICA.

1 Introduction

Embeddings play an important role in natural language processing, ranging from word embeddings (Mikolov et al., 2013) to internal representations in language models (Devlin et al., 2019; Brown et al., 2020; Touvron et al., 2023). Understanding how embeddings represent meaning is crucial for unraveling black box NLP models.

Independent Component Analysis (ICA) (Hyvärinen and Oja, 2000) is an effective method for visualizing and interpreting the geometric structure of embeddings (Musil and Mareček, 2024; Yamagiwa et al., 2023). Just as PCA aims to make coordinate axes uncorrelated, ICA seeks to transform the coordinate axes into statistically independent components. The resulting axes from ICA tend to have sparser component values with a few larger values compared to PCA, which increases interpretability as the axes can be seen as specific semantic components (Fig. 1).

However it has been pointed out that the estimated 'independent components' are only approximately independent (Hyvärinen et al., 2001; Sasaki et al., 2013, 2014). This is because many real-world datasets cannot be accurately represented as a linear combination of independent components, contradicting the assumption of ICA theory.

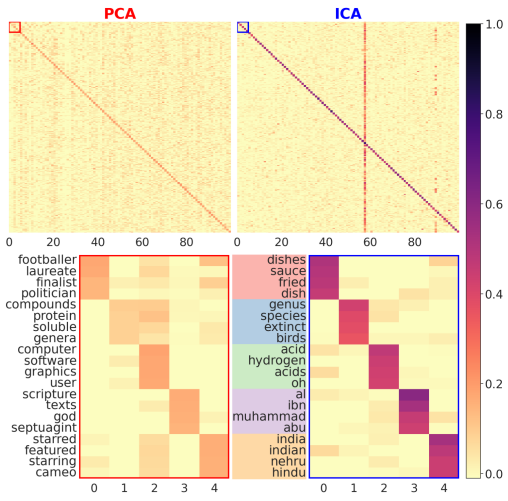


Figure 1: Heatmap visualization of 300-dimensional SGNS embeddings transformed by PCA and ICA, with axes sorted by variance and skewness, respectively. Each embedding has been normalized to have a norm of 1 for better visual interpretation. For each axis, the top 4 words (frequency $n_w \geq 100$ in text8) with largest component values were used. The first 100 axes are displayed in the top panels, and the first 5 axes with the word labels are displayed in the bottom panels. See Appendices A, B and G for details.

In this study, we aim to further interpret the results of applying ICA to embeddings by focusing on the non-independence between 'independent components'. We measure the degree of non-independence by calculating higher-order correlations between components and find that components with large higher-order correlations can be interpreted as having strong semantic associations. The entire structure is revealed by visualizing the maximum spanning tree of semantic components with higher-order correlations as edge weights.

Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pages 2883–2899

2 Review: ICA-Transformed Embeddings

Procedure of ICA. For a centered embedding matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ that represents the meanings of n words by d-dimensional vectors, ICA[1] seeks a transformation $\mathbf{S} = \mathbf{XB}$ such that each component $S_1, \cdots, S_d$ of the transformed matrix $\mathbf{S} = [S_1, \cdots, S_d]$ is as statistically independent as possible[2]. The transformation $\mathbf{B}$ can be expressed as the product of the whitening transformation matrix $\mathbf{A}$ (e.g., PCA transformation) and the orthogonal transformation matrix $\mathbf{R}_{\text{ica}}$, i.e., the resulting $\mathbf{S}$ is represented as

$$\mathbf{S} = \mathbf{XAR}_{\text{ica}}. \tag{1}$$

Here, $\mathbf{R}_{\text{ica}}$ is obtained by minimizing the mutual information[3] $I(S_1 \cdots S_d) = \sum H(S_i) - H(S_1 \cdots S_d)$, which is equivalent to maximizing the non-gaussianity[4] of the distributions of S_i (Hyvärinen and Oja, 2000). The normalized ICA-transformed embeddings, with each embedding in $\mathbf{S}$ rescaled to a norm of 1, offer high interpretability (Yamagiwa et al., 2023, 2024) and are used for visualizations in this paper.

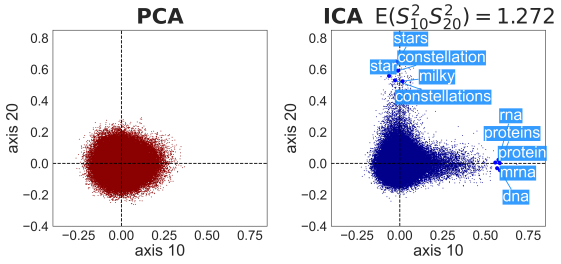


Figure 2: Scatterplots of normalized word embeddings along the 10th and 20th axes. The axes for PCA and ICA-transformed embeddings were arranged in descending order of variance and skewness, respectively. In both transformations, the components are uncorrelated.

Comparison of PCA and ICA. Figure 2 shows that ICA can find the 'spiky and interpretable shape' of the embedding distribution (e.g., "biology" and "stars" for the 10th and 20th axes, respectively), but PCA cannot. This is because ICA determines the coordinate axes toward high non-gaussianity, while PCA only considers variance information.

[1]For the computation, `FastICA` (Hyvärinen, 1999) implemented in `scikit-learn` (Pedregosa et al., 2011) is used.

[2]The k-th component S_k is also referred to as Axis k.

[3]$H(X) = -\int P_X(x) \log P_X(x) dx$ is the entropy.

[4]The degree to which a probability distribution deviates from a Gaussian distribution can be measured using statistics based on higher-order moments, such as skewness (the third moment) or the negentropy of the distribution.

3 Higher-Order Correlations Among Estimated Independent Components

Non-Independence in Real-World Data. The 'independent components' estimated by ICA on real-world data are uncorrelated but not completely independent, with dependencies existing between components (Hyvärinen et al., 2001; Sasaki et al., 2013, 2014). This is because ICA assumes a linear decomposition into independent components, an assumption frequently violated in reality.

Higher-Order Correlation. To quantify non-independencies, methods like mutual information and Hilbert-Schmidt Independence Criterion (HSIC) (Gretton et al., 2005) exist. Here we use the higher-order correlation, the simplest measure in terms of computation and formulation. This measure is expressed as follows:

$$\mathrm{E}(S_i^2 S_j^2) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{S}_{t,i}^2 \mathbf{S}_{t,j}^2. \tag{2}$$

Here, $\mathbf{S}$ is the whitened matrix[5]. This can also be interpreted as the covariance between S_i^2 and S_j^2, plus one, as $\mathrm{cov}(S_i^2, S_j^2) = \mathrm{E}((S_i^2 - 1)(S_j^2 - 1)) = \mathrm{E}(S_i^2 S_j^2) - 1$. If S_i and S_j are independent of each other, then $\mathrm{E}(S_i^2 S_j^2) = \mathrm{E}(S_i^2)\mathrm{E}(S_j^2) = 1$. Thus, the deviation of $\mathrm{E}(S_i^2 S_j^2)$ from 1 is the degree of dependence between S_i and S_j.

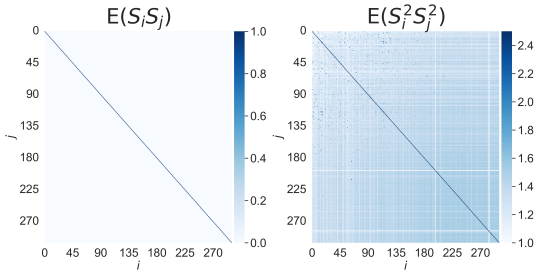


Figure 3: Heatmaps of the correlation coefficient $\mathrm{E}(S_i S_j)$ and the higher-order correlation $\mathrm{E}(S_i^2 S_j^2)$ of component pairs (S_i, S_j) from ICA on 300-dimensional SGNS embeddings. See Appendix C for details.

Figure 3 shows that the estimated independent components of the embeddings are uncorrelated but not completely independent, with varying degrees of higher-order correlations across pairs. These $\mathrm{E}(S_i^2 S_j^2)$ values provide a useful metric of association, as demonstrated in the following section.

[5]The components are (i) centered: $\mathrm{E}(S_i) = 0$, the mean of each component is 0, (ii) scaled: $\mathrm{E}(S_i^2) = 1$, the variance of each component is 1, and (iii) uncorrelated: $\mathrm{E}(S_i S_j) = 0$, the correlations are all zero.

$\mathrm{E}(S_0^2 S_{82}^2) = 1.927$		$\mathrm{E}(S_6^2 S_{96}^2) = 2.032$		$\mathrm{E}(S_{12}^2 S_{66}^2) = 1.975$		$\mathrm{E}(S_{16}^2 S_{118}^2) = 2.124$		$\mathrm{E}(S_{56}^2 S_{126}^2) = 1.861$		$\mathrm{E}(S_{63}^2 S_{210}^2) = 2.964$	
Axis 0	Axis 82	Axis 6	Axis 96	Axis 12	Axis 66	Axis 16	Axis 118	Axis 56	Axis 126	Axis 63	Axis 210
dishes	beer	el	o	rabbi	judah	blood	disorder	cpu	pointer	organization	unesco
sauce	beers	spanish	portuguese	talmud	israelites	organs	mental	microprocessor	return	international	itu
fried	ale	nacional	paulo	rabbis	yahweh	liver	disorders	processor	string	organizations	interpol
dish	brewing	jos	rio	torah	elisha	kidney	symptoms	cpus	pointers	interpol	observer
cooked	yeast	de	portugal	jewish	isaiah	tissue	bipolar	intel	node	standardization	temporary
$\mathrm{E}(S_0^2 S_{23}^2) = 0.990$		$\mathrm{E}(S_6^2 S_{13}^2) = 0.992$		$\mathrm{E}(S_{12}^2 S_{57}^2) = 0.993$		$\mathrm{E}(S_{16}^2 S_{57}^2) = 0.996$		$\mathrm{E}(S_{56}^2 S_{197}^2) = 0.982$		$\mathrm{E}(S_{63}^2 S_{18}^2) = 1.073$	
Axis 0	Axis 23	Axis 6	Axis 13	Axis 12	Axis 57	Axis 16	Axis 57	Axis 56	Axis 197	Axis 63	Axis 18
dishes	statesman	el	windows	rabbi	s	blood	s	cpu	population	organization	actress
sauce	astronomer	spanish	os	talmud	and	organs	and	microprocessor	median	international	footballer
fried	philosopher	nacional	unix	rabbis	was	liver	was	processor	estimated	organizations	musician
dish	johann	jos	linux	torah	in	kidney	in	cpus	residing	interpol	actor
cooked	mathematician	de	microsoft	jewish	by	tissue	by	intel	total	standardization	singer

Table 1: (Top Row) Six randomly selected pairs of components from the top 50 pairs with the highest $|\mathrm{E}(S_i^2 S_j^2) - 1|$ values. For each component, the top 5 words (frequency $n_w \geq 100$ in text8) with the largest component values are listed. (Bottom Row) Component pairs with small $\mathrm{E}(S_i^2 S_k^2)$ values. For each component S_i with the smaller axis number in the pairs (S_i, S_j) in the top row, a component S_k with the smallest value of $|\mathrm{E}(S_i^2 S_k^2) - 1|$ was selected.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
List-2 (top-k)	**69.0**	**65.0**	**64.0**	**64.5**	**56.5**
List-3 (bottom 30%)	27.0	33.0	32.5	33.5	40.5
Can't decide	4.0	2.0	3.5	2.0	3.0

Table 2: The percentage of each list judged by the GPT model to be more semantically related to List-1.

4 Interpretation of Higher-Order Correlations as Semantic Relevance

4.1 Degree of Semantic Relevance

We show that the values of higher-order correlations $\mathrm{E}(S_i^2 S_j^2)$ can be interpreted as the degree of associations between semantic components.

Results: Top Row of Table 1. The meanings of each component, represented by the listed words in component pairs with high $\mathrm{E}(S_i^2 S_j^2)$ values, are strongly related. For example, focusing on Axis 0 and Axis 82, a pair with particularly large values of $\mathrm{E}(S_i^2 S_j^2)$, we can interpret that Axis 0 has a meaning associated with "dishes" and Axis 82 with "beer", suggesting that there is a semantic relationship between them.

Results: Bottom Row of Table 1. On the other hand, for component pairs with $\mathrm{E}(S_i^2 S_j^2)$ values close to 1, indicating that the components are considered independent, there is no clear relevance between the meanings of the components. For example, looking at the pair of Axis 0 and Axis 23, which has a small $\mathrm{E}(S_i^2 S_j^2)$ value, we can interpret that Axis 0 represents "dishes" and Axis 23 represents "polymath", and there is no direct semantic relationship between them.

Detailed results are shown in Appendix G.

4.2 Quantitative Evaluation via GPT-4o mini

We conducted experiments to quantitatively evaluate whether higher-order correlations between ICA components represent semantic relationships.

Settings. Our experimental procedure was as follows. We first selected the top 100 ICA components, ranked by skewness. For each component i ($i = 0, \ldots, 99$), we created three word lists: Word list-1 comprised the top 5 words from component i, Word list-2 contained the top 5 words from the k-th most correlated component with component i ($k = 1, \ldots, 5$), and Word list-3 consisted of the top 5 words from a randomly selected low-correlation component (chosen from the bottom 30% of correlated components). Using these lists, we generated pairs (list-1, list-2) and (list-1, list-3), and queried GPT-4o mini to determine which pair was more semantically related[6]. This procedure was executed for all 100 components, resulting in 200 total comparisons for each value of k from 1 to 5. The specific prompt used for GPT-4o mini model is provided in Appendix D.

Results and Discussion. Table 2 shows the result of the experiment. We can see that component pairs with higher-order correlations tend to be more semantically related (69.0% for $k = 1$ vs 27.0% for bottom 30%), and that semantic relatedness gradually declines as correlation decreases (69.0% at $k = 1$ to 56.5% at $k = 5$). These results quantitatively demonstrate that higher-order correlations between ICA components effectively reflect semantic relatedness between corresponding words.

[6]To mitigate potential biases, we randomly shuffled the order of the lists in the pairs and repeated this process with the reversed order: (list-1, list-3) and (list-1, list-2).

$\mathrm{E}(S^2_{10}S^2_2) = 2.323$		$\mathrm{E}(S^2_{10}S^2_{16}) = 1.947$		$\mathrm{E}(S^2_{10}S^2_{160}) = 1.811$		$\mathrm{E}(S^2_{27}S^2_{11}) = 1.643$		$\mathrm{E}(S^2_{27}S^2_{64}) = 1.997$		$\mathrm{E}(S^2_{27}S^2_{104}) = 1.605$	
Axis 10	Axis 2	Axis 10	Axis 16	Axis 10	Axis 160	Axis 27	Axis 11	Axis 27	Axis 64	Axis 27	Axis 104
dna	acid	dna	blood	dna	evolution	greek	gaius	greek	goddess	greek	archaeological
proteins	hydrogen	proteins	organs	proteins	evolutionary	greece	caesar	greece	gods	greece	neolithic
rna	acids	rna	liver	rna	darwin	athens	augustus	athens	deity	athens	bc
mrna	oh	mrna	kidney	mrna	selection	athenian	lucius	athenian	deities	athenian	pottery
w_k	$\mathbf{S}^2_{k,10}\mathbf{S}^2_{k,2}$	w_k	$\mathbf{S}^2_{k,10}\mathbf{S}^2_{k,16}$	w_k	$\mathbf{S}^2_{k,10}\mathbf{S}^2_{k,160}$	w_k	$\mathbf{S}^2_{k,27}\mathbf{S}^2_{k,11}$	w_k	$\mathbf{S}^2_{k,27}\mathbf{S}^2_{k,64}$	w_k	$\mathbf{S}^2_{k,27}\mathbf{S}^2_{k,104}$
ribose	3755.7	adenylate	2079.8	utr	2381.5	laertius	898.3	demeter	2348.5	tiryns	1690.6
deoxyribose	2963.9	effectors	1842.5	reticulum	1942.0	preveza	788.0	hephaestus	2204.3	knossos	1348.1
phosphodiester	2850.2	antisense	1639.9	genomic	1668.6	xanthippus	764.2	hestia	2021.5	mycenaean	1205.6
biosynthesis	2510.1	cyclase	1638.9	homozygous	1599.1	rhadamanthus	735.5	hera	1744.6	lendering	1124.7
methyltransferase	2482.9	myosin	1201.8	cleaved	1181.0	thracians	711.8	cronos	1720.2	hissarlik	1103.1
pyrimidine	2399.6	axons	1144.2	tubulin	1152.4	alexandri	705.2	aphrodite	1675.9	melos	1006.3

Table 3: For 6 component pairs (S_i, S_j) selected from adjacent component pairs in the MST defined in Sec. 5, the top 6 words and their corresponding $\mathbf{S}^2_{t,i}\mathbf{S}^2_{t,j}$ values that contribute the most to the $\mathrm{E}(S^2_i S^2_j)$ value are shown.

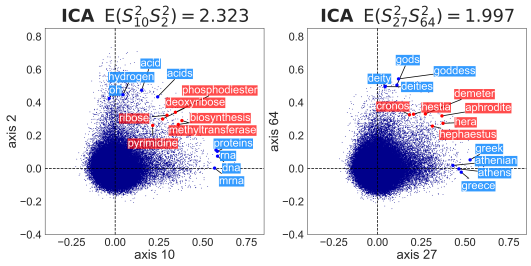


Figure 4: Scatter plots of normalized word embeddings for axis pairs (10, 2) and (27, 64) with large values of higher-order correlations. Blue-labeled words are the top 4 words for each axis's component values, while red-labeled words are the top 6 words for the values of $\mathbf{S}^2_{t,i}\mathbf{S}^2_{t,j}$. See Appendix C for all the pairs in Table 3.

4.3 Decomposition of Semantic Relevance

For a component pair (S_i, S_j), words w_t with large values of $\mathbf{S}^2_{t,i}\mathbf{S}^2_{t,j}$ are considered to make a significant contribution to the higher-order correlation $\mathrm{E}(S^2_i S^2_j) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{S}^2_{t,i}\mathbf{S}^2_{t,j}$. Here we investigate words that significantly contribute to the $\mathrm{E}(S^2_i S^2_j)$ values and gain a more concrete understanding of the relationships between components.

Results. Table 3 presents component pairs selected from the maximum spanning tree T_{150} (Sec. 5) and words significantly contributing to their $\mathrm{E}(S^2_i S^2_j)$ values. These words often relate to the meanings of both components, demonstrating additive compositionality. For example, in the Axis 10 and Axis 2 pair, words like *ribose*, *deoxyribose*, *phosphodiester*, *biosynthesis*, *methyltransferase*, and *pyrimidine* notably contribute to the $\mathrm{E}(S^2_i S^2_j)$ value, linking biomolecules and chemical components. Detailed results are shown in Appendix G.

Visualization. Figure 4 shows word embedding scatter plots for axis pairs (10, 2) and (27, 64) with large higher-order correlations to illustrate the distribution of words with significant contributions to higher-order correlations. Unlike a typical independent component scatter plot (Fig. 2), these exhibit many words with large component values in both axes, reflecting the meanings of both axes and demonstrating the additive compositionality of embeddings. For the (10, 2) pair, words that notably contribute to the $\mathrm{E}(S^2_i S^2_j)$ (*ribose*, *deoxyribose*, *phosphodiester*, *biosynthesis*, *methyltransferase*, and *pyrimidine*) appear with significant values in both components. This abundance of words sharing both semantic components is characteristic of pairs with large higher-order correlations. Detailed results are shown in Appendix C.

5 Visualization of Non-Independence Structure

In this section, we construct a maximum spanning tree (MST) based on higher-order correlations to visualize the non-independence between estimated independent components.

Settings. The 300 ICA components, originally ordered by skewness with $i = 0, \cdots, 299$, were re-sorted in descending order of semantic component consistency scores to prioritize axes that are more easily interpretable as specific semantic components. The semantic component consistency scores were determined by a word intrusion task (Chang et al., 2009). A higher consistency score indicates easier interpretability. Details of the scoring methods are provided in Appendix E.1. We introduce the notation σ to map the order of consistency scores to the original axis numbers in the skewness order: $\sigma(j)$ represents the axis number in the skewness sort for the axis with the j-th highest consistency score. Then, we consider a weighted complete graph G_{150}, with 150 components having high consistency scores $S_{\sigma(i)}$ $(i \in 0, \cdots, 149)$ as nodes. For the edge between the node pair (S_i, S_j),

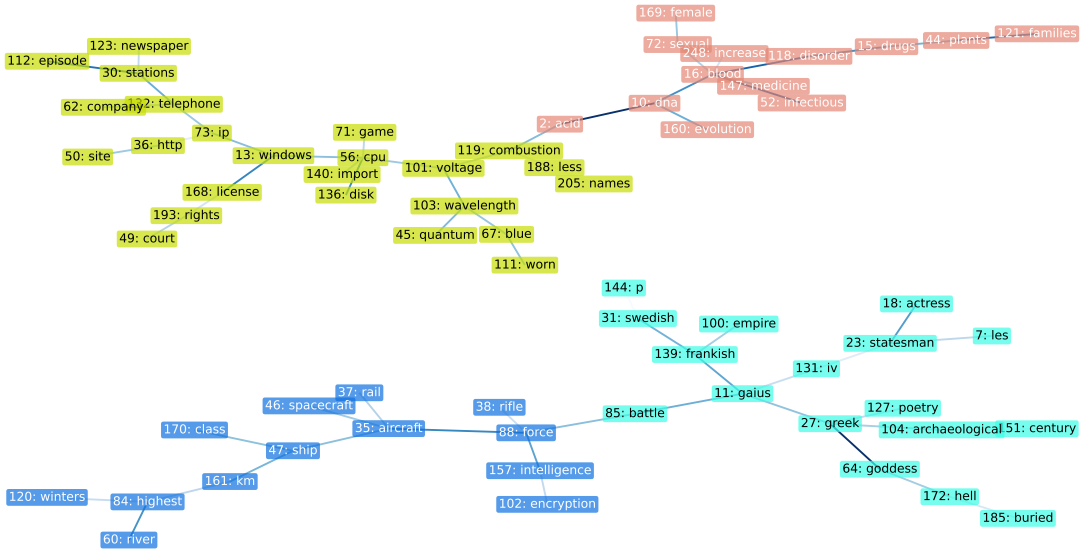


Figure 5: Subtrees of MST T_{150} defined in Sec. 5. Each node represents an independent component S_k (i.e., Axis k) estimated by ICA. The label of each node is "k : TopWord(k)", where TopWord(k) is the word with the largest component value along axis k among words with frequency $n_w \geq 100$ in the text8 corpus. The color of the edge between nodes (i, j) represents the magnitude of the $\mathrm{E}(S_i^2 S_j^2)$ value between the components, with darker edge colors indicating larger values.

we set $c_{ij} = \mathrm{E}(S_i^2 S_j^2)$ as the weight. To visualize and interpret G_{150}, we compute the maximum spanning tree (MST)[7] T_{150}, a spanning tree that maximizes the sum of c_{ij} in the graph G_{150}. MST was relatively more interpretable than other subgraphs of graph G_{150}, providing a good balance between visibility and element relationships.

Interpretation of the MST. The MST T_{150} represents a graph structure expressing the non-independence between estimated independent components. Since the edges in T_{150} connect component pairs with large higher-order correlations, we can interpret that there is a strong relationship between the components connected by these edges. Furthermore, the subtrees of T_{150} represent groups of semantically related components, and the components within these groups tend to have similar meanings.

Results and Discussion. Figure 5 shows a part of the MST T_{150}; the entire MST T_{150} is exhibited in Appendix E.2. The colors correspond to the clusters obtained by applying spectral clustering[8] (Ng et al., 2001) to T_{150}. The weights used for clustering are the higher-order correlations. From the MST, we can infer structures such as connections and groupings of meanings among three or more components[9]. For example, semantically related component pairs such as (*2: dna, 10: acid*) in the pink cluster and (*27: greek, 64: goddess*) in the cyan cluster are connected by edges in T_{150}. Additionally, groups such as {*168: license, 13: windows, 56: cpu*} in the yellow cluster and {*46: spacecraft, 35: aircraft, 47: ship*} in the blue cluster form semantic clusters as sets of nodes connected by edges. The components within these groups can be interpreted as having meanings related to "computer" and "vehicle", respectively.

6 Conclusion

Both ICA and PCA transformations make the components uncorrelated. ICA goes further by making the components nearly independent, but some non-independence still remains. In this study, we used higher-order correlations to quantify the non-independence between the components in the ICA-transformed word embeddings. By interpreting these as the semantic associations between the components and visualizing the overall structure, we can gain a deeper understanding of the latent semantic structure within the embeddings.

[7] We used `minimum_spanning_tree` implemented in `NetworkX` (Hagberg et al., 2008) for the computation of the MST. See Appendix E.2 for details.

[8] We used `SpectralClustering` implemented in `scikit-learn` (Pedregosa et al., 2011).

[9] Furthermore, in Appendix F, we conducted a dimensionality reduction experiment that numerically demonstrates how the MST effectively represents a significant structure among the components.

Limitations

- The embeddings used in the experiments are limited to SGNS word embeddings. For a more thorough analysis, it is necessary to conduct experiments using various types of embeddings.
- For large embedding matrices with a high number of data points n, ICA may fail to converge within a practical timeframe. To overcome this, we suggest using subsampled data to estimate the ICA transformation matrix, which can then be applied to unseen embedding vectors.
- When n is large, calculating higher-order correlations (eq. 2) may become computationally intensive. This calculation is similar to the computation of the variance-covariance matrix and can be approximated by subsampling data points. Further speedup can be achieved by parallelization of the computation.
- Since ICA leverages the non-Gaussianity of embedding distributions, it is not suitable for analysis if the original embeddings follow a multivariate Gaussian distribution.

Ethics Statement

This study complies with the ACL Ethics Policy.

Acknowledgements

This study was partially supported by JSPS KAKENHI 22H05106, 23H03355, JST CREST JPMJCR21N3, JST BOOST JPMJBS2407, JST SPRING JPMJSP2110.

Code Availability

Code is available at `https://github.com/momoseoyama/hoc`.

References

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS*, pages 1877–1901.

Elia Bruni, Nam-Khanh Tran, and Marco Baroni. 2014. Multimodal distributional semantics. *Journal of Artificial Intelligence Research*, 49:1–47.

Jonathan Chang, Sean Gerrish, Chong Wang, Jordan Boyd-graber, and David Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Lev Finkelstein, Evgeniy Gabrilovich, Yossi Matias, Ehud Rivlin, Zach Solan, Gadi Wolfman, and Eytan Ruppin. 2002. Placing search in context: The concept revisited. *ACM Transactions on information systems*, 20(1):116–131.

Daniela Gerz, Ivan Vulić, Felix Hill, Roi Reichart, and Anna Korhonen. 2016. SimVerb-3500: A large-scale evaluation set of verb similarity. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2173–2182.

Arthur Gretton, Olivier Bousquet, Alex Smola, and Schölkopf Bernhard. 2005. Measuring statistical dependence with hilbert-schmidt norms. In *The International Conference on Algorithmic Learning Theory, ALT*, pages 63–77.

Aric A Hagberg, Daniel A Schult, and Pieter J Swart. 2008. Exploring network structure, dynamics, and function using networkx. *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, pages 11–15.

Felix Hill, Roi Reichart, and Anna Korhonen. 2015. Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, 41(4):665–695.

Aapo Hyvärinen. 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634.

Aapo Hyvärinen, Patrik O. Hoyer, and Mika Inki. 2001. Topographic independent component analysis. *Neural Computation*, 13(7):1527–1558.

Aapo Hyvärinen and Erkki Oja. 2000. Independent component analysis: Algorithms and applications. *Neural networks*, 13(4-5):411–430.

Thang Luong, Richard Socher, and Christopher Manning. 2013. Better word representations with recursive neural networks for morphology. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 104–113.

Matt Mahoney. 2011. About the test data. http://mattmahoney.net/dc/textdata.html.

Tomás Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems26: Annual Conference on Neural Information Processing Systems 2013, NeurIPS*, pages 3111–3119.

Tomáš Musil and David Mareček. 2024. Exploring interpretability of independent components of word embeddings with automated word intruder test. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 6922–6928, Torino, Italia. ELRA and ICCL.

Andrew Ng, Michael Jordan, and Weiss. 2001. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems14: Annual Conference on Neural Information Processing Systems 2001, NIPS*.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830.

Kira Radinsky, Eugene Agichtein, Evgeniy Gabrilovich, and Shaul Markovitch. 2011. A word at a time: Computing word relatedness using temporal semantic analysis. In *Proceedings of the 20th International Conference on World Wide Web*, page 337–346.

Hiroaki Sasaki, Michael Gutmann, Hayaru Shouno, and Aapo Hyvärinen. 2013. Correlated topographic analysis: estimating an ordering of correlated components. *Machine Learning*, 92:285–317.

Hiroaki Sasaki, Michael Gutmann, Hayaru Shouno, and Aapo Hyvärinen. 2014. Estimating Dependency Structures for non-Gaussian Components with Linear and Energy Correlations. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, AISTATS*, pages 868–876.

Fei Sun, J. Guo, Yanyan Lan, Jun Xu, and Xueqi Cheng. 2016. Sparse word embeddings using $l1$ regularized online learning. In *International Joint Conference on Artificial Intelligence*.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Hiroaki Yamagiwa, Momose Oyama, and Hidetoshi Shimodaira. 2023. Discovering universal geometry in embeddings with ICA. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 4647–4675, Singapore. Association for Computational Linguistics.

Hiroaki Yamagiwa, Momose Oyama, and Hidetoshi Shimodaira. 2024. Revisiting cosine similarity via normalized ica-transformed embeddings. *arXiv preprint arXiv:2406.10984*.

A Details of Experimental Settings

The word embeddings used in the experiments were trained using Skip-gram with Negative Sampling (SGNS). The parameters used to train SGNS are summarized in Table 4. The corpus used for training is the text8 corpus (Mahoney, 2011), and the number of vocabulary words is $n = 253{,}854$.

Dimensionality	300
Epochs	100
Window size h	10
Negative samples ν	5
Learning rate	0.025
Min count	1

Table 4: SGNS parameters.

B Remarks on Axis 57 in Figure 1

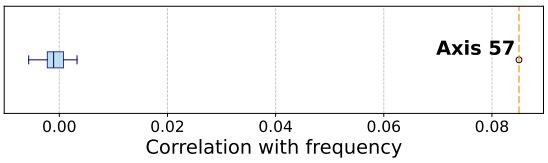


Figure 6: Boxplot of correlation coefficients between word frequency n_w and the component values for the 0th to 99th axes of the ICA-transformed embeddings. Axis 57 shows a particularly high correlation coefficient.

An interesting vertical streak is observed in Axis 57 of the heatmap for ICA-transformed embeddings in Fig. 1. This streak can be explained by several factors. As shown in Fig. 6, Axis 57 exhibits a strong correlation between component values and word frequencies n_w, suggesting that Axis 57 is more associated with word frequency than with a specific semantic meaning. Additionally, the words used in Fig. 1 were selected from those appearing more than 100 times in the text8 corpus, resulting in a bias towards high-frequency

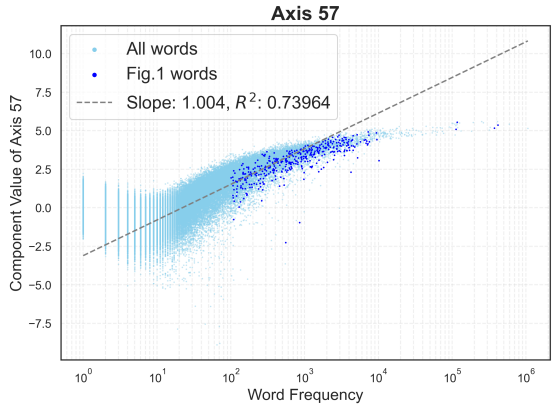


Figure 7: Scatter plot of word frequency n_w versus the component values of the 57th axis of the ICA-transformed embeddings. Words used in Fig. 1 are highlighted in dark blue. The regression line and coefficient of determination were calculated for words with a frequency of $n_w \geq 10$.

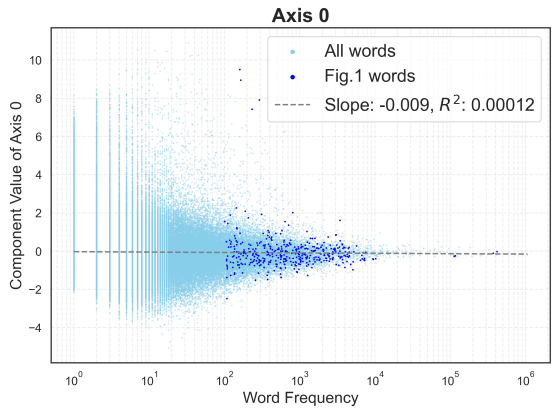


Figure 8: Scatter plot of word frequency n_w versus the component values of the 0th axis of the ICA-transformed embeddings. The settings are the same as in Fig. 7.

words. This tendency is further illustrated in Fig. 7, which demonstrates that words used in the heatmap ($n_w \geq 100$) tend to have larger component values along Axis 57. In contrast, Fig. 8 shows that for axes with weak correlation to word frequency, the words used in the heatmap do not exhibit notably large component values. Consequently, large component values were observed along Axis 57 in the heatmap, a pattern that was unique to Axis 57 and not observed in other axes.

C Higher-Order Correlations

C.1 Distribution of Higher-Order Correlations

Figures 9a and 9b show histograms of higher-order correlations $\mathrm{E}(S_i^2 S_j^2)$ for pairs where $i < j$ and for all pairs including $\mathrm{E}(S_i^4)$ where $i = j$, respectively. While there are component pairs where $\mathrm{E}(S_i^2 S_j^2) < 1$, in Fig. 3, the range of values was truncated between 1.0 and 2.5 for visualization purposes.

C.2 Scatterplots for Independent Axes

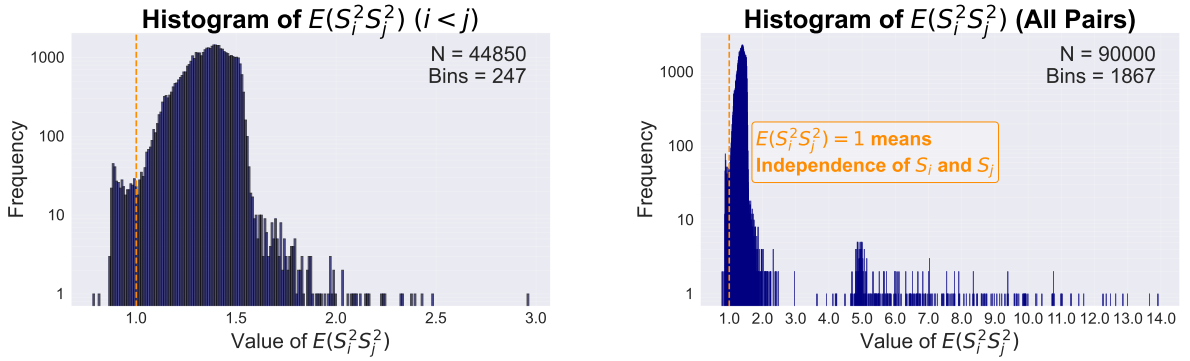
Complementary Results for Sec. 4.3. Table 3 in Sec. 4.3 presented words with significant contributions to higher-order correlations for six axis pairs. While the main text illustrated the distribution of these highly contributing words through scatter plots for the two selected pairs, Figure 10 provides scatter plots for all the six pairs.

The Relationship Between the Magnitude of Higher-Order Correlations and the Appearance of Scatter Plots. Figure 11 presents the scatter plots of word embeddings for 24 component pairs, each with different higher-order correlation values. We can see that as the magnitude of higher-order correlation increases, the number of words with large component values along both axes increases as well. The selection of these 24 pairs was conducted as follows: First, we considered 150 components $S_{\sigma(0)}, \cdots, S_{\sigma(149)}$ with high semantic consistency (see Appendix E.1 for the calculation method). We then sorted all possible component pairs (S_i, S_j) $(i, j \in \sigma(0), \cdots, \sigma(149))$ based on the value of $|\mathrm{E}(S_i^2 S_j^2) - 1|$. We established 24 equally spaced grids between the minimum and maximum values, and selected pairs closest to each grid point without repetition.

Words with Significant Contributions to Higher-Order Correlations. We have observed words with significant contributions, i.e., with large values of $\mathbf{S}_{t,i}^2 \mathbf{S}_{t,j}^2$, to higher-order correlations $\mathrm{E}(S_i^2 S_j^2)$ in Table 3 in Sec. 4.3, and will see further examples in Tables 8 and 9 in Appendix G. Such words are labeled in red in the scatter plots in Fig. 4 in Section 4.3, as well as in Figs. 10 and 11 in Appendix C.2. For axis pairs with large higher-order correlations, we observe a large number of words that make significant contributions to the higher-order correlations. The meanings of these words include both axes' meanings, demonstrating the additive compositionality of embeddings.

D Prompt Used for Evaluation by GPT-4o mini

The specific prompt used for the GPT-4o mini model is provided below.



(a) For (i, j) pairs where $i < j$.

(b) For all (i, j) pairs, including $i = j$.

Figure 9: Histograms of higher-order correlations.

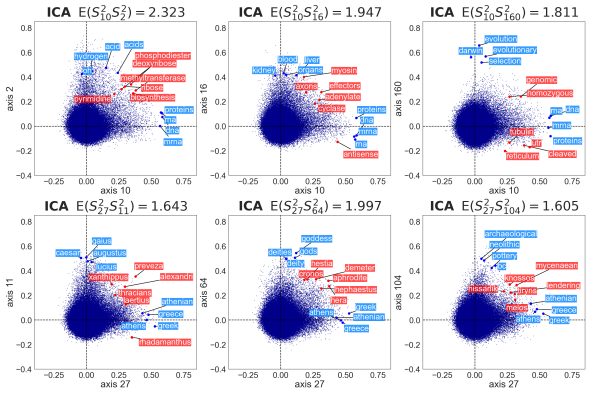


Figure 10: Scatter plots of normalized word embeddings for axis pairs in Table 3. Blue-labeled words are the top 4 words for each axis's component values, while red-labeled words are the top 6 words for values of $\mathbf{S}_{t,i}^2\mathbf{S}_{t,j}^2$.

```
Question:
  You are given 2 list pairs (A, B), (C, D).
  If one pair is more semantically relevant than
      the other, answer the pair.
  If you cannot determine, answer "XX".

List pair (A, B): ({wordlist_1}, {wordlist_2})
List pair (C, D): ({wordlist_1}, {wordlist_3})

Output:
  "AB" if (A, B) is more semantically related
  "CD" if (C, D) is more semantically related
  "XX" if equally related, or you can't decide
Respond with only AB, CD, or XX.
```

When conducting the experiment, we took the following steps to eliminate any potential biases arising from the order of word sequences and specific label names.

1. To remove the influence of word order within lists, we randomly shuffled the words in `wordlist_1`, `wordlist_2`, and `wordlist_3`.

2. To prevent bias in output labels, we conducted the experiment twice, swapping `wordlist_2` and `wordlist_3` between runs.

3. To account for order bias in the prompt, we randomly alternated the order of the following two lines:

```
List pair (A, B): ({wordlist_1}, {wordlist_2})
List pair (C, D): ({wordlist_1}, {wordlist_3})
```

E Details of Visualization of Non-Independence Structure

E.1 Scoring ICA Axes: Word Intrusion Task

We assigned a semantic coherence score to each axis of the ICA-transformed embeddings using the word intrusion task method (Chang et al., 2009).

Word Intrusion Task. The word intrusion task is a method used to evaluate the semantic coherence of a set of k words by assessing the ability to identify an intruder word. For instance, consider the set of words {*windows, os, unix, linux, microsoft*}, which has a consistent theme of operating systems. In this case, an unrelated word such as *waterskiing* should be easily identifiable as an intruder, as it does not align with the theme of operating systems. In our experiment, we assigned coherence scores to the top $k = 5$ words (with frequency $n_w \geq 100$ in the text8 corpus) for each axis.

Selection of the Intruder Word. In order to select the intruder word for the set of top k words of each axis $a \in \{1, \ldots, d\}$, denoted as $\text{top}_k(a)$, we randomly chose a word from a pool of words that satisfy both of the following criteria simultaneously: (i) the word ranks in the lower 50% in terms of the component value on the axis a, and (ii) it ranks in the top 10% in terms of the component value on some axis other than a. For each axis, $L = 100$ intruder words are randomly selected, and $W_{\text{int}}(a)$ denotes the set of these L intruder words.

Scoring Method. For the consistency score of the meaning of each axis a, $\mathrm{Score}(a)$, we adopted the metric proposed by Sun et al. (2016).

$$\mathrm{Score}(a) = \frac{\mathrm{InterDist}(a)}{\mathrm{IntraDist}(a)}$$

$$\mathrm{IntraDist}(a) = \sum_{\substack{w_i, w_j \in \mathrm{top}_k(a) \\ w_i \neq w_j}} \frac{\mathrm{dist}(w_i, w_j)}{k(k-1)}$$

$$\mathrm{InterDist}(a) = \underset{w \in W_{\mathrm{int}}(a)}{\mathrm{mean}} \sum_{w_i \in \mathrm{top}_k(a)} \frac{\mathrm{dist}(w_i, w)}{k}$$

In this formula, we defined $\mathrm{dist}(w_i, w_j) = \|\mathbf{s}_i - \mathbf{s}_j\|$ for the ICA-transformed embeddings. Here, $\mathrm{IntraDist}(a)$ denotes the average distance between the top k words, and $\mathrm{InterDist}(a)$ represents the average distance between the top words and the intruder words. The score is higher when the intruder words are further away from the set $\mathrm{top}_k(a)$. Therefore, this score serves as a quantitative measure of the ability to identify the intruder word, thus it is used as a measure of the consistency of the meaning of the top k words and the interpretability of axes.

E.2 Entire Visualization of MST

Figure 12 is the visualization of maximum spanning tree (MST) T_{150} defined in Sec. 5. For a graph G_{150}, where the cost between nodes i and j defined as $c_{ij} = E(S_i^2 S_j^2)$, the algorithm to find the MST T is a greedy method that maximizes the total sum of costs, $\sum_{(i,j)\in T} c_{ij}$, subject to T being a spanning tree. The greedy algorithm selects edges in decreasing order of c_{ij} while adhering to the tree constraint. Due to the monotonicity of $f(x) = 1/x$, the decreasing order of c_{ij} is equivalent to the increasing order of $1/c_{ij}$. Thus, computing the MST T_{150} in the graph G_{150} is equivalent to finding the minimum spanning tree, which minimizes the sum of $1/c_{ij}$.

F Dimensionality Reduction via MST

	$d=2$	$d=5$	$d=10$	$d=20$	$d=50$	$d=100$
Random Clustering on components (PCA)	0.04	0.08	0.12	0.16	0.23	0.29
Random Clustering on components (ICA)	0.04	0.08	0.12	0.17	0.24	0.29
Spectral Clustering on MST (PCA)	0.03	0.08	0.13	0.17	0.24	0.30
Spectral Clustering on MST (ICA)	**0.06**	**0.13**	**0.18**	**0.23**	**0.28**	**0.31**

Table 5: Word similarity scores for dimensionality reduction.

We experimentally confirmed that the structure of higher-order correlations between components can be applied to the dimensionality reduction of embeddings. Specifically, we performed Spectral Clustering on the maximum Spanning Tree (MST) T_{300}, which was computed based on the higher-order correlations between components. By reducing the dimensionality of the embeddings through averaging the clustered axes, we verified that the accuracy degradation in Word Similarity Tasks was mitigated compared to random clustering.

Experimental Settings. We conducted our experiments using the 300-dimensional word embeddings (SGNS). These embeddings were subjected to PCA and ICA to obtain components for clustering. We employed two clustering methods: (1) Random Clustering and (2) Spectral Clustering on the maximum Spanning Tree (MST) T_{300}, which was computed based on the higher-order correlations calculated using Eq. 2. Clustering was performed with the number of clusters ranging from 2 to 100. Dimensionality reduction was achieved by averaging the clustered axes, resulting in reduced dimensions from $d = 2$ to $d = 100$. The performance of the lower-dimensional embeddings was evaluated through Word Similarity Tasks. For the Word Similarity Tasks, we utilized six datasets: MEN (Bruni et al., 2014), WS353 (Finkelstein et al., 2002), MTurk (Radinsky et al., 2011), RW (Luong et al., 2013), SimLex999 (Hill et al., 2015), and SimVerb-3500 (Gerz et al., 2016). Each dataset comprises word pairs along with gold similarity scores, assigned by human annotators. We employed the Spearman rank correlation coefficient between human ratings and the cosine similarity of the word embeddings as the evaluation metric. The reported values represent the average scores across the six datasets.

Results and Discussion The experimental results are presented in Table 5. We observe that ICA-based methods outperform PCA-based methods. Moreover, our proposed method, Spectral Clustering on the MST of ICA components, consistently achieves the best performance across all dimensions. This can be attributed to the fact that components included in the same cluster on the MST likely have high semantic relevance and play similar roles in representing the meaning of words. These results validate that considering higher-order correlations between axes better preserves semantic relationships in compressed word embeddings, demonstrating the practical utility of our method. It is important to note that this evaluation assesses the

performance of clustering using a downstream task of dimensionality reduction, rather than dimensionality reduction itself.

G Supplementary Tables for ICA Components and MST Subtrees

Table 6 shows all components of the ICA-transformed word embeddings used in our experiments. Associated with the experiments in Sec. 4.1, Table 7 shows the top 60 pairs with the highest $\mathrm{E}(S_i^2 S_j^2)$ values. Additionally, in Table 8 and Table 9, we report on all component pairs in the subtrees of MST T_{150} shown in Fig. 5, extending the results from the selected pairs previously reported in Sec. 4.3.

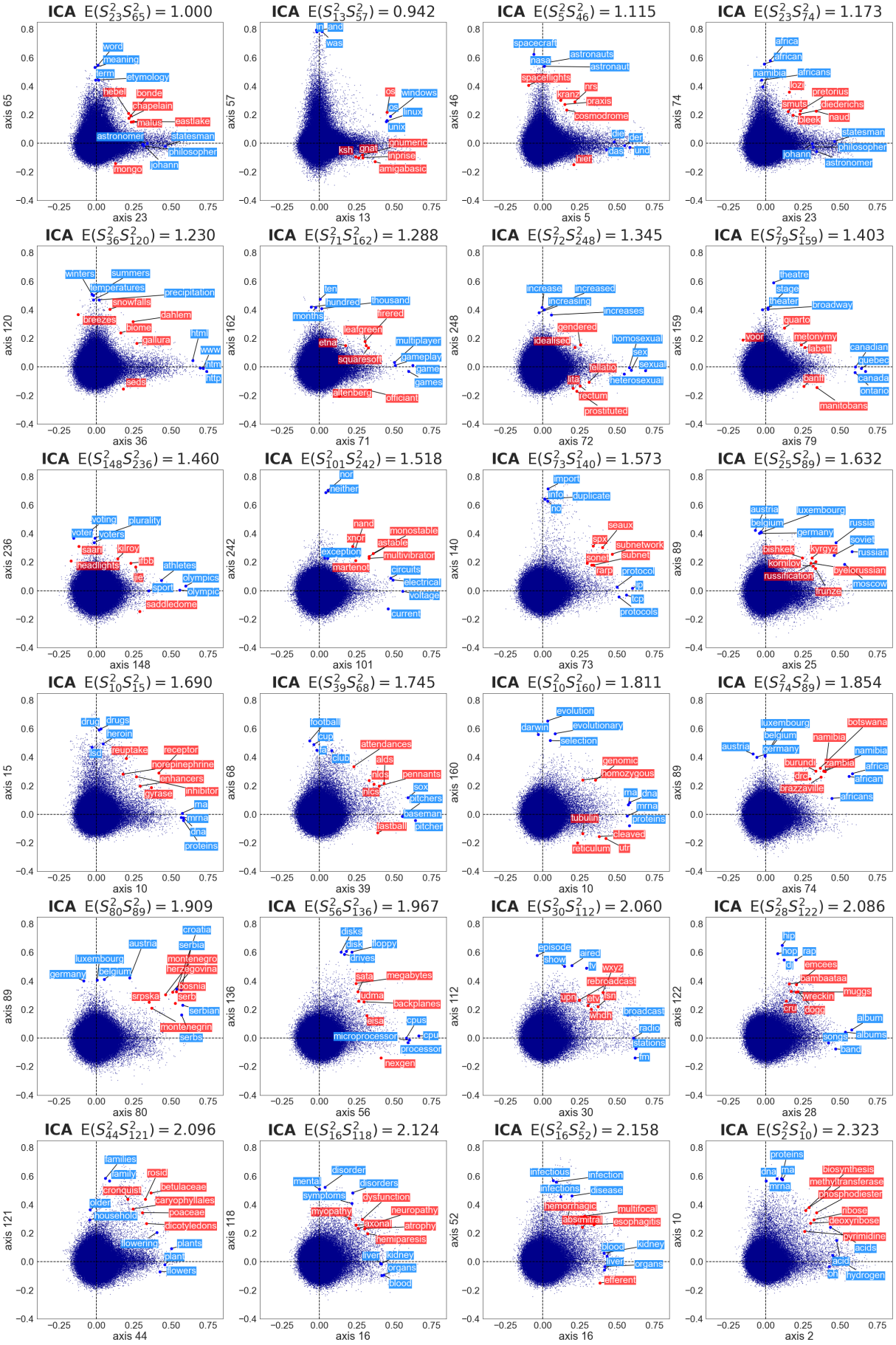


Figure 11: Scatter plots of normalized word embeddings for 24 axis pairs. Blue-labeled words are the top 4 words for each axis's component values, while red-labeled words are the top 6 words for values of $\mathbf{S}^2_{t,i}\mathbf{S}^2_{t,j}$.

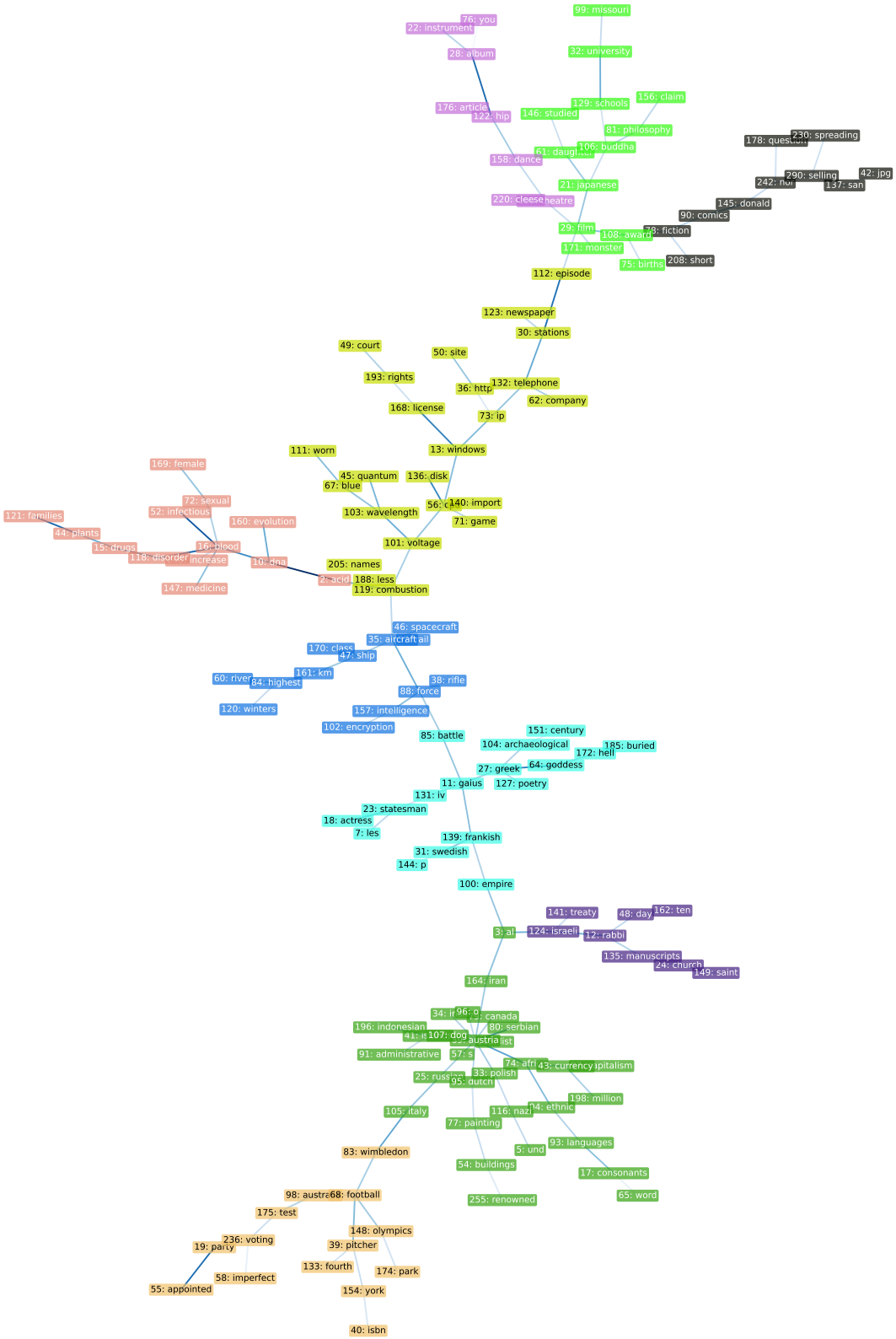


Figure 12: Visualization of the entire MST T_{150} defined in Sec. 5. Each color of the nodes represents one of the ten clusters obtained by Spectral Clustering.

Axis 0	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7	Axis 8	Axis 9	Axis 10	Axis 11	Axis 12	Axis 13	Axis 14
dishes	genus	acid	al	india	und	el	les	chinese	di	dna	gaius	rabbi	windows	topological
sauce	species	hydrogen	ibn	indian	der	spanish	du	china	italian	proteins	caesar	talmud	os	isomorphic
fried	extinct	acids	muhammad	nehru	die	nacional	des	beijing	della	rna	augustus	rabbis	unix	banach
dish	birds	oh	abu	hindu	das	jos	paris	pinyin	luigi	mrna	lucius	torah	linux	topology
Axis 15	Axis 16	Axis 17	Axis 18	Axis 19	Axis 20	Axis 21	Axis 22	Axis 23	Axis 24	Axis 25	Axis 26	Axis 27	Axis 28	Axis 29
drugs	blood	consonants	actress	party	stars	japanese	instrument	statesman	church	russian	cars	greek	album	film
drug	organs	vowels	footballer	parties	constellation	japan	instruments	astronomer	churches	moscow	ford	greece	albums	films
heroin	liver	vowel	musician	democrats	star	tokyo	bass	philosopher	communion	russia	car	athens	band	directed
lsd	kidney	consonant	actor	democratic	constellations	emperor	guitars	johann	orthodox	soviet	chassis	athenian	songs	director
Axis 30	Axis 31	Axis 32	Axis 33	Axis 34	Axis 35	Axis 36	Axis 37	Axis 38	Axis 39	Axis 40	Axis 41	Axis 42	Axis 43	Axis 44
stations	swedish	university	polish	irish	aircraft	http	rail	rifle	pitcher	isbn	islands	jpg	currency	plants
fm	sweden	college	poland	ireland	flight	www	trains	gun	sox	press	island	image	currencies	plant
radio	danish	technology	krak	dublin	boeing	htm	railway	rifles	baseman	ed	archipelago	png	euro	flowers
broadcast	norwegian	institute	aw	ulster	airlines	html	train	ammunition	pitchers	routledge	atoll	gif	dollar	flowering
Axis 45	Axis 46	Axis 47	Axis 48	Axis 49	Axis 50	Axis 51	Axis 52	Axis 53	Axis 54	Axis 55	Axis 56	Axis 57	Axis 58	Axis 59
quantum	spacecraft	ship	day	court	site	coach	infectious	element	buildings	appointed	cpu	s	imperfect	concerto
particles	nasa	ships	observances	judge	website	quarterback	infection	metals	building	minister	microprocessor	and	perfect	fugue
particle	astronauts	hms	holidays	courts	forum	defensive	disease	elements	tower	cabinet	processor	was	future	sonata
physics	astronaut	cruisers	holiday	trial	photos	bengals	infections	uranium	built	appoints	cpus	in	present	bwv
Axis 60	Axis 61	Axis 62	Axis 63	Axis 64	Axis 65	Axis 66	Axis 67	Axis 68	Axis 69	Axis 70	Axis 71	Axis 72	Axis 73	Axis 74
river	daughter	company	organization	goddess	word	judah	blue	football	gesserit	accusative	game	sexual	ip	africa
tributaries	married	corporation	international	gods	meaning	israelites	white	cup	bene	nouns	games	sex	tcp	african
rivers	marriage	companies	organizations	deity	etymology	yahweh	red	fa	leto	genitive	gameplay	homosexual	protocols	africans
navigable	wife	shareholders	interpol	deities	term	elisha	yellow	club	duncan	noun	multiplayer	heterosexual	protocol	namibia
Axis 75	Axis 76	Axis 77	Axis 78	Axis 79	Axis 80	Axis 81	Axis 82	Axis 83	Axis 84	Axis 85	Axis 86	Axis 87	Axis 88	Axis 89
births	you	painting	fiction	canada	serbian	philosophy	beer	wimbledon	highest	battle	rocks	solids	force	austria
deaths	know	paintings	novels	quebec	serbs	philosophical	beers	open	elevation	battles	volcanic	faces	army	belgium
alumni	me	art	novel	canadian	serbia	kant	ale	finalist	ft	defeat	granite	vertices	military	luxembourg
novelists	we	painters	stories	ontario	croatia	philosophers	brewing	quarter	lowest	fought	geologic	cube	regiment	germany
Axis 90	Axis 91	Axis 92	Axis 93	Axis 94	Axis 95	Axis 96	Axis 97	Axis 98	Axis 99	Axis 100	Axis 101	Axis 102	Axis 103	Axis 104
comics	administrative	essex	languages	ethnic	dutch	o	capitalism	australia	missouri	empire	voltage	encryption	wavelength	archaeological
marvel	divided	somerset	spoken	peoples	van	portuguese	anarcho	australian	kentucky	empires	electrical	cryptography	light	neolithic
superhero	districts	cornwall	language	indigenous	netherlands	paulo	capitalists	sydney	mississippi	emperors	circuits	cipher	wavelengths	bc
superman	divisions	exeter	dialects	minorities	amsterdam	rio	economists	melbourne	alabama	rulers	current	ciphers	laser	pottery
Axis 105	Axis 106	Axis 107	Axis 108	Axis 109	Axis 110	Axis 111	Axis 112	Axis 113	Axis 114	Axis 115	Axis 116	Axis 117	Axis 118	Axis 119
italy	buddha	dog	award	computation	martial	worn	episode	frac	morphisms	audio	nazi	si	disorder	combustion
norway	buddhism	hound	awards	artificial	judo	wearing	aired	cos	homomorphism	lossy	holocaust	units	mental	diesel
netherlands	mahayana	dogs	prize	turing	aikido	clothing	show	equation	hydrogen	compression	nazis	metre	disorders	turbine
germany	buddhist	breed	awarded	computational	karate	wear	tv	euler	wavelengths	mpeg	camps	kilogram	symptoms	engine
Axis 120	Axis 121	Axis 122	Axis 123	Axis 124	Axis 125	Axis 126	Axis 127	Axis 128	Axis 129	Axis 130	Axis 131	Axis 132	Axis 133	Axis 134
winters	families	hip	newspaper	israeli	algae	pointer	poetry	ball	schools	exports	iv	telephone	fourth	prix
summers	family	hop	daily	palestinian	bacteria	return	verse	scrimmage	school	imports	iii	phone	third	grand
temperatures	older	dj	weekly	palestinians	fungi	string	poem	goal	secondary	textiles	vii	mobile	fifth	schumacher
precipitation	household	rap	newspapers	israel	mitochondria	pointers	poems	foul	education	gwh	vi	cellular	sixth	race
Axis 135	Axis 136	Axis 137	Axis 138	Axis 139	Axis 140	Axis 141	Axis 142	Axis 143	Axis 144	Axis 145	Axis 146	Axis 147	Axis 148	Axis 149
manuscripts	disk	san	horse	frankish	import	treaty	colspan	card	p	donald	studied	medicine	olympics	saint
translation	floppy	francisco	horses	franks	duplicate	signed	align	cards	j	duck	born	medical	olympic	st
translations	disks	california	riding	saxon	info	agreement	center	dealer	q	scrooge	he	doctors	athletes	petersburg
testament	drives	diego	breed	saxons	no	signing	motto	betting	r	mcduck	career	care	sport	patron
Axis 150	Axis 151	Axis 152	Axis 153	Axis 154	Axis 155	Axis 156	Axis 157	Axis 158	Axis 159	Axis 160	Axis 161	Axis 162	Axis 163	Axis 164
nuclear	century	inducted	finns	york	list	claim	intelligence	dance	theatre	evolution	km	ten	corpus	iran
bomb	th	treasure	soviets	new	topics	claims	agency	dances	stage	evolutionary	harbors	hundred	vitamin	iranian
bombs	twentieth	apocalypse	col	ny	lists	claimed	cia	dancing	broadway	darwin	total	months	esp	iraq
fission	nineteenth	fairy	ir	bronx	see	evidence	fbi	dancers	theater	selection	unpaved	thousand	mutant	persian
Axis 165	Axis 166	Axis 167	Axis 168	Axis 169	Axis 170	Axis 171	Axis 172	Axis 173	Axis 174	Axis 175	Axis 176	Axis 177	Axis 178	Axis 179
desertification	axis	thank	license	female	class	monster	hell	basque	park	test	article	arthur	question	bah
environment	perpendicular	hello	copyleft	male	classes	creature	heaven	spain	parks	tests	main	merlin	debate	ll
flooding	rotation	please	gpl	age	middle	loch	paradise	aragon	national	testing	disambiguation	grail	questions	mason
environmental	direction	good	licenses	infant	working	creatures	eden	eta	wildlife	cricket	discusses	knights	whether	meher
Axis 180	Axis 181	Axis 182	Axis 183	Axis 184	Axis 185	Axis 186	Axis 187	Axis 188	Axis 189	Axis 190	Axis 191	Axis 192	Axis 193	Axis 194
jedi	coup	ethiopia	turkish	g	buried	reading	cort	less	taggart	diamond	melville	guant	rights	freemasonry
luke	overthrow	ethiopian	turkey	e	burial	further	georges	than	pg	diamonds	fuller	namo	legislation	masonic
knight	tat	eritrea	istanbul	icao	cemetery	devised	philip	more	estonian	stones	apr	consistency	act	dawn
wars	junta	kate	cypriot	fao	grave	steadily	counts	much	finnish	gem	taste	alter	laws	lodge
Axis 195	Axis 196	Axis 197	Axis 198	Axis 199	Axis 200	Axis 201	Axis 202	Axis 203	Axis 204	Axis 205	Axis 206	Axis 207	Axis 208	Axis 209
cretaceous	indonesian	population	million	derivative	columbus	actor	cross	centered	maps	names	rolling	chemistry	short	code
geologic	malaysia	median	estimated	manchester	taxi	squadron	crescent	eagles	map	various	adobe	alchemy	long	calling
epoch	indonesia	estimated	billion	resolutions	benson	actress	icrc	serial	gis	include	chuck	tubes	lived	codes
extinction	malaysian	residing	estimates	pat	concern	isps	dressing	eventual	gps	these	kasparov	orientation	conventional	anthem
Axis 210	Axis 211	Axis 212	Axis 213	Axis 214	Axis 215	Axis 216	Axis 217	Axis 218	Axis 219	Axis 220	Axis 221	Axis 222	Axis 223	Axis 224
unesco	calvinism	k	z	tyler	tex	franklin	esperanto	tended	foo	cleese	certificate	oed	anonymous	caves
itu	laden	defunct	rich	cornell	marsh	heath	ido	dave	slight	monty	employee	hungry	gap	alone
interpol	trotsky	neutrality	zwingli	edit	extermination	falklands	mi	emissions	needle	python	engineer	angry	collections	elsewhere
observer	limited	ln	swiss	theorem	iliad	boas	ne	assist	folklore	palin	laplace	threats	incidence	daniel
Axis 225	Axis 226	Axis 227	Axis 228	Axis 229	Axis 230	Axis 231	Axis 232	Axis 233	Axis 234	Axis 235	Axis 236	Axis 237	Axis 238	Axis 239
lord	lenin	principia	pray	bees	spreading	et	aunt	hong	oswald	pm	voting	mad	martin	pierre
rings	brands	mainframe	erasmus	bee	giving	hoc	blade	kong	bull	tunnel	plurality	distinguishing	luther	lynx
admiralty	conan	running	raises	ants	popularity	ad	loan	guidance	op	sec	voter	classify	jr	khmer
courtesy	russell	stroke	facing	ant	expertise	est	accidental	mine	distinct	privileges	voters	karen	zwingli	bases
Axis 240	Axis 241	Axis 242	Axis 243	Axis 244	Axis 245	Axis 246	Axis 247	Axis 248	Axis 249	Axis 250	Axis 251	Axis 252	Axis 253	Axis 254
albert	aether	nor	wood	prevalence	wayne	guant	sentences	increase	handling	explosives	ultimate	joins	jack	advocacy
einstein	accelerated	neither	arrows	worst	ds	namo	lebesgue	increased	replication	convenience	reduction	comparatively	amos	professor
piercing	populace	anything	aspartame	cutting	exhibits	rn	neil	increasing	oliver	eccentric	discover	breed	plug	pirate
combining	bear	exception	merge	bluetooth	ears	holes	sigma	increases	ignore	quasi	striking	fungi	doug	cuban
Axis 255	Axis 256	Axis 257	Axis 258	Axis 259	Axis 260	Axis 261	Axis 262	Axis 263	Axis 264	Axis 265	Axis 266	Axis 267	Axis 268	Axis 269
renowned	pickford	humanities	fatal	attending	devastating	cheeses	accessed	farmer	metaphor	landmarks	glider	textbook	definitive	criticized
tourists	coins	statement	provided	xvi	prints	witchcraft	arbor	clown	sanctioned	cage	trademark	looks	continually	mead
home	turbulent	conversation	tcp	fate	assertion	cheese	ann	desk	strictly	br	purity	springer	eukaryotic	stressed
town	earn	witches	descendant	quad	ink	oath	fahrenheit	metaphysical	sagan	ruin	flew	seals	gradient	classified
Axis 270	Axis 271	Axis 272	Axis 273	Axis 274	Axis 275	Axis 276	Axis 277	Axis 278	Axis 279	Axis 280	Axis 281	Axis 282	Axis 283	Axis 284
postage	doraemon	importantly	dodo	crusader	napoleon	plots	blake	adjacent	constructs	feces	cryonics	cosmology	economical	multiply
partly	adaptation	crusade	girlfriend	ceremony	bonaparte	hired	ne	proposed	hierarchy	buying	ec	reunited	comprised	guernsey
willing	inuit	juan	punk	mailing	herman	twisted	cap	machines	synod	chess	euthanasia	ll	initiative	contraception
binary	bones	carpenter	mcgill	euclidean	reed	characterize	chances	ban	ois	shadows	freezing	cosmological	trap	dependency
Axis 285	Axis 286	Axis 287	Axis 288	Axis 289	Axis 290	Axis 291	Axis 292	Axis 293	Axis 294	Axis 295	Axis 296	Axis 297	Axis 298	Axis 299
union	wiki	lynch	mortal	tributaries	selling	treated	agave	countryside	beckham	fitness	fragile	self	cumberland	gardner
unrelated	communities	turned	assassin	sources	copies	desk	expense	accession	analyses	glass	rousseau	prince	dylan	correction
lifelong	usenet	ordered	fortified	hormones	aquatic	sullivan	audience	burn	pictures	regular	clever	vested	functions	habsburg
arm	lightning	meals	dmt	helicopters	sell	tables	var	completion	trinidad	marilyn	planet	ordained	kay	codes

Table 6: The top 4 words with the largest component values along all axes of ICA-transformed word embedding used in our experiments. Axes are sorted in descending order of skewness.

$E(S^2_{63}S^2_{210}) = 2.964$		$E(S^2_{2}S^2_{114}) = 2.480$		$E(S^2_{10}S^2_{125}) = 2.431$		$E(S^2_{58}S^2_{70}) = 2.395$		$E(S^2_{14}S^2_{113}) = 2.380$		$E(S^2_{63}S^2_{184}) = 2.330$	
Axis 63	Axis 210	Axis 2	Axis 114	Axis 10	Axis 125	Axis 58	Axis 70	Axis 14	Axis 113	Axis 63	Axis 184
organization	unesco	acid	morphisms	dna	algae	imperfect	accusative	topological	frac	organization	g
international	itu	hydrogen	homomorphism	proteins	bacteria	perfect	nouns	isomorphic	cos	international	e
organizations	interpol	acids	hydrogen	rna	fungi	future	genitive	banach	equation	organizations	icao
interpol	observer	oh	wavelengths	mrna	mitochondria	present	noun	topology	euler	interpol	fao
$E(S^2_{2}S^2_{10}) = 2.323$		$E(S^2_{22}S^2_{59}) = 2.247$		$E(S^2_{26}S^2_{134}) = 2.233$		$E(S^2_{14}S^2_{114}) = 2.228$		$E(S^2_{2}S^2_{53}) = 2.165$		$E(S^2_{16}S^2_{52}) = 2.158$	
Axis 2	Axis 10	Axis 22	Axis 59	Axis 26	Axis 134	Axis 14	Axis 114	Axis 2	Axis 53	Axis 16	Axis 52
acid	dna	instrument	concerto	cars	prix	topological	morphisms	acid	element	blood	infectious
hydrogen	proteins	instruments	fugue	ford	grand	isomorphic	homomorphism	hydrogen	metals	organs	infection
acids	rna	bass	sonata	car	schumacher	banach	hydrogen	acids	elements	liver	disease
oh	mrna	guitars	bwv	chassis	race	topology	wavelengths	oh	uranium	kidney	infections
$E(S^2_{10}S^2_{114}) = 2.150$		$E(S^2_{16}S^2_{118}) = 2.124$		$E(S^2_{1}S^2_{121}) = 2.107$		$E(S^2_{44}S^2_{121}) = 2.096$		$E(S^2_{28}S^2_{122}) = 2.086$		$E(S^2_{19}S^2_{55}) = 2.075$	
Axis 10	Axis 114	Axis 16	Axis 118	Axis 1	Axis 121	Axis 44	Axis 121	Axis 28	Axis 122	Axis 19	Axis 55
dna	morphisms	blood	disorder	genus	families	plants	families	album	hip	party	appointed
proteins	homomorphism	organs	mental	species	family	plant	family	albums	hop	parties	minister
rna	hydrogen	liver	disorders	extinct	older	flowers	older	band	dj	democrats	cabinet
mrna	wavelengths	kidney	symptoms	birds	household	flowering	household	songs	rap	democratic	appoints
$E(S^2_{30}S^2_{112}) = 2.060$		$E(S^2_{6}S^2_{96}) = 2.032$		$E(S^2_{21}S^2_{110}) = 2.028$		$E(S^2_{27}S^2_{64}) = 1.997$		$E(S^2_{13}S^2_{168}) = 1.991$		$E(S^2_{51}S^2_{128}) = 1.987$	
Axis 30	Axis 112	Axis 6	Axis 96	Axis 21	Axis 110	Axis 27	Axis 64	Axis 13	Axis 168	Axis 51	Axis 128
stations	episode	el	o	japanese	martial	greek	goddess	windows	license	coach	ball
fm	aired	spanish	portuguese	japan	judo	greece	gods	os	copyleft	quarterback	scrimmage
radio	show	nacional	paulo	tokyo	aikido	athens	deity	unix	gpl	defensive	goal
broadcast	tv	jos	rio	emperor	karate	athenian	deities	linux	licenses	bengals	foul
$E(S^2_{12}S^2_{66}) = 1.975$		$E(S^2_{4}S^2_{106}) = 1.974$		$E(S^2_{56}S^2_{136}) = 1.967$		$E(S^2_{17}S^2_{114}) = 1.966$		$E(S^2_{15}S^2_{118}) = 1.959$		$E(S^2_{10}S^2_{16}) = 1.947$	
Axis 12	Axis 66	Axis 4	Axis 106	Axis 56	Axis 136	Axis 17	Axis 114	Axis 15	Axis 118	Axis 10	Axis 16
rabbi	judah	india	buddha	cpu	disk	consonants	morphisms	drugs	disorder	dna	blood
talmud	israelites	indian	buddhism	microprocessor	floppy	vowels	homomorphism	drug	mental	proteins	organs
rabbis	yahweh	nehru	mahayana	processor	disks	vowel	hydrogen	heroin	disorders	rna	liver
torah	elisha	hindu	buddhist	cpus	drives	consonant	wavelengths	lsd	symptoms	mrna	kidney
$E(S^2_{0}S^2_{82}) = 1.927$		$E(S^2_{53}S^2_{150}) = 1.915$		$E(S^2_{80}S^2_{89}) = 1.909$		$E(S^2_{5}S^2_{59}) = 1.897$		$E(S^2_{39}S^2_{51}) = 1.885$		$E(S^2_{10}S^2_{52}) = 1.875$	
Axis 0	Axis 82	Axis 53	Axis 150	Axis 80	Axis 89	Axis 5	Axis 59	Axis 39	Axis 51	Axis 10	Axis 52
dishes	beer	element	nuclear	serbian	austria	und	concerto	pitcher	coach	dna	infectious
sauce	beers	metals	bomb	serbs	belgium	der	fugue	sox	quarterback	proteins	infection
fried	ale	elements	bombs	serbia	luxembourg	die	sonata	baseman	defensive	rna	disease
dish	brewing	uranium	fission	croatia	germany	das	bwv	pitchers	bengals	mrna	infections
$E(S^2_{53}S^2_{86}) = 1.872$		$E(S^2_{39}S^2_{128}) = 1.871$		$E(S^2_{6}S^2_{173}) = 1.869$		$E(S^2_{8}S^2_{106}) = 1.867$		$E(S^2_{56}S^2_{126}) = 1.861$		$E(S^2_{85}S^2_{153}) = 1.857$	
Axis 53	Axis 86	Axis 39	Axis 128	Axis 6	Axis 173	Axis 8	Axis 106	Axis 56	Axis 126	Axis 85	Axis 153
element	rocks	pitcher	ball	el	basque	chinese	buddha	cpu	pointer	battle	finns
metals	volcanic	sox	scrimmage	spanish	spain	china	buddhism	microprocessor	return	battles	soviets
elements	granite	baseman	goal	nacional	aragon	beijing	mahayana	processor	string	defeat	col
uranium	geologic	pitchers	foul	jos	eta	pinyin	buddhist	cpus	pointers	fought	ir
$E(S^2_{74}S^2_{89}) = 1.854$		$E(S^2_{15}S^2_{16}) = 1.834$		$E(S^2_{83}S^2_{105}) = 1.832$		$E(S^2_{17}S^2_{233}) = 1.828$		$E(S^2_{71}S^2_{143}) = 1.828$		$E(S^2_{53}S^2_{114}) = 1.824$	
Axis 74	Axis 89	Axis 15	Axis 16	Axis 83	Axis 105	Axis 17	Axis 233	Axis 71	Axis 143	Axis 53	Axis 114
africa	austria	drugs	blood	wimbledon	italy	consonants	hong	game	card	element	morphisms
african	belgium	drug	organs	open	norway	vowels	kong	games	cards	metals	homomorphism
africans	luxembourg	heroin	liver	finalist	netherlands	vowel	guidance	gameplay	dealer	elements	hydrogen
namibia	germany	lsd	kidney	quarter	germany	consonant	mine	multiplayer	betting	uranium	wavelengths
$E(S^2_{30}S^2_{132}) = 1.820$		$E(S^2_{10}S^2_{160}) = 1.811$		$E(S^2_{86}S^2_{195}) = 1.809$		$E(S^2_{101}S^2_{103}) = 1.800$		$E(S^2_{35}S^2_{88}) = 1.798$		$E(S^2_{4}S^2_{89}) = 1.794$	
Axis 30	Axis 132	Axis 10	Axis 160	Axis 86	Axis 195	Axis 101	Axis 103	Axis 35	Axis 88	Axis 4	Axis 89
stations	telephone	dna	evolution	rocks	cretaceous	voltage	wavelength	aircraft	force	india	austria
fm	phone	proteins	evolutionary	volcanic	geologic	electrical	light	flight	army	indian	belgium
radio	mobile	rna	darwin	granite	epoch	circuits	wavelengths	boeing	military	nehru	luxembourg
broadcast	cellular	mrna	selection	geologic	extinction	current	laser	airlines	regiment	hindu	germany
$E(S^2_{107}S^2_{138}) = 1.793$		$E(S^2_{15}S^2_{44}) = 1.793$		$E(S^2_{32}S^2_{129}) = 1.793$		$E(S^2_{20}S^2_{46}) = 1.788$		$E(S^2_{3}S^2_{124}) = 1.786$		$E(S^2_{26}S^2_{119}) = 1.785$	
Axis 107	Axis 138	Axis 15	Axis 44	Axis 32	Axis 129	Axis 20	Axis 46	Axis 3	Axis 124	Axis 26	Axis 119
dog	horse	drugs	plants	university	schools	stars	spacecraft	al	israeli	cars	combustion
hound	horses	drug	plant	college	school	constellation	nasa	ibn	palestinian	ford	diesel
dogs	riding	heroin	flowers	technology	secondary	star	astronauts	muhammad	palestinians	car	turbine
breed	breed	lsd	flowering	institute	education	constellations	astronaut	abu	israel	chassis	engine

Table 7: Complementary experimental results to Table 1. The top 60 pairs with the highest $E(S_i^2 S_j^2)$ values are presented. For each component, the top 4 words with the largest component values are listed.

$\mathrm{E}(S_2^2 S_{10}^2) = 2.323$		$\mathrm{E}(S_2^2 S_{119}^2) = 1.755$		$\mathrm{E}(S_{132}^2 S_{30}^2) = 1.820$		$\mathrm{E}(S_{132}^2 S_{73}^2) = 1.693$		$\mathrm{E}(S_{132}^2 S_{62}^2) = 1.632$		$\mathrm{E}(S_{136}^2 S_{56}^2) = 1.967$	
Axis 2	Axis 10	Axis 2	Axis 119	Axis 132	Axis 30	Axis 132	Axis 73	Axis 132	Axis 62	Axis 136	Axis 56
acid	dna	acid	combustion	telephone	stations	telephone	ip	telephone	company	disk	cpu
hydrogen	proteins	hydrogen	diesel	phone	fm	phone	tcp	phone	corporation	floppy	microprocessor
acids	rna	acids	turbine	mobile	radio	mobile	protocols	mobile	companies	disks	processor
oh	mrna	oh	engine	cellular	broadcast	cellular	protocol	cellular	shareholders	drives	cpus
w_k	$\mathbf{S}_{k,2}^2 \mathbf{S}_{k,10}^2$	w_k	$\mathbf{S}_{k,2}^2 \mathbf{S}_{k,119}^2$	w_k	$\mathbf{S}_{k,132}^2 \mathbf{S}_{k,30}^2$	w_k	$\mathbf{S}_{k,132}^2 \mathbf{S}_{k,73}^2$	w_k	$\mathbf{S}_{k,132}^2 \mathbf{S}_{k,62}^2$	w_k	$\mathbf{S}_{k,136}^2 \mathbf{S}_{k,56}^2$
ribose	3755.7	pyrolysis	2794.4	digitalized	4726.8	multipoint	2062.7	esat	3145.5	sata	3427.5
deoxyribose	2963.9	syngas	2056.9	arabsat	4657.2	pstn	1996.9	telecoms	2547.9	udma	2519.7
phosphodiester	2850.2	gasification	1783.5	radiotelephone	4453.9	wimax	1873.8	nynex	2155.9	backplanes	2008.9
biosynthesis	2510.1	butane	1761.2	landlines	3522.1	xdsl	1491.2	gnc	1810.0	nexgen	1947.7
methyltransferase	2482.9	dehydrogenation	1623.0	intersputnik	3053.5	svcs	1361.2	haitel	1657.3	megabytes	1890.1
pyrimidine	2399.6	tert	1230.2	telex	2722.4	isdn	1235.8	openreach	1529.2	eisa	1859.5
$\mathrm{E}(S_{10}^2 S_{16}^2) = 1.947$		$\mathrm{E}(S_{10}^2 S_{160}^2) = 1.811$		$\mathrm{E}(S_{140}^2 S_{56}^2) = 1.615$		$\mathrm{E}(S_{13}^2 S_{56}^2) = 1.732$		$\mathrm{E}(S_{13}^2 S_{168}^2) = 1.991$		$\mathrm{E}(S_{13}^2 S_{73}^2) = 1.740$	
Axis 10	Axis 16	Axis 10	Axis 160	Axis 140	Axis 56	Axis 13	Axis 56	Axis 13	Axis 168	Axis 13	Axis 73
dna	blood	dna	evolution	import	cpu	windows	cpu	windows	license	windows	ip
proteins	organs	proteins	evolutionary	duplicate	microprocessor	os	microprocessor	os	copyleft	os	tcp
rna	liver	rna	darwin	info	processor	unix	processor	unix	gpl	unix	protocols
mrna	kidney	mrna	selection	no	cpus	linux	cpus	linux	licenses	linux	protocol
w_k	$\mathbf{S}_{k,10}^2 \mathbf{S}_{k,16}^2$	w_k	$\mathbf{S}_{k,10}^2 \mathbf{S}_{k,160}^2$	w_k	$\mathbf{S}_{k,140}^2 \mathbf{S}_{k,56}^2$	w_k	$\mathbf{S}_{k,13}^2 \mathbf{S}_{k,56}^2$	w_k	$\mathbf{S}_{k,13}^2 \mathbf{S}_{k,168}^2$	w_k	$\mathbf{S}_{k,13}^2 \mathbf{S}_{k,73}^2$
adenylate	2079.8	utr	2381.5	superpipelined	5652.9	xcode	2609.7	qpl	5678.2	netware	1799.2
effectors	1842.5	reticulum	1942.0	strongarm	3220.2	powerpc	2046.7	lgpl	4519.9	netbios	1543.2
antisense	1639.9	genomic	1668.6	specrate	2470.5	itanium	1500.1	trolltech	3588.4	imap	1414.0
cyclase	1638.9	homozygous	1599.1	specbaserate	1524.5	glibc	1355.1	gpl	3325.2	glut	1239.0
myosin	1201.8	cleaved	1181.0	insubstantial	1387.8	irix	1177.1	gnu	2826.1	wfw	1179.9
axons	1144.2	tubulin	1152.4	eisa	1027.8	efi	1161.1	bsd	2822.7	dhcpv	1115.5
$\mathrm{E}(S_{15}^2 S_{118}^2) = 1.959$		$\mathrm{E}(S_{15}^2 S_{44}^2) = 1.793$		$\mathrm{E}(S_{16}^2 S_{248}^2) = 1.499$		$\mathrm{E}(S_{16}^2 S_{147}^2) = 1.717$		$\mathrm{E}(S_{16}^2 S_{72}^2) = 1.628$		$\mathrm{E}(S_{16}^2 S_{52}^2) = 2.158$	
Axis 15	Axis 118	Axis 15	Axis 44	Axis 16	Axis 248	Axis 16	Axis 147	Axis 16	Axis 72	Axis 16	Axis 52
drugs	disorder	drugs	plants	blood	increase	blood	medicine	blood	sexual	blood	infectious
drug	mental	drug	plant	organs	increased	organs	medical	organs	sex	organs	infection
heroin	disorders	heroin	flowers	liver	increasing	liver	doctors	liver	homosexual	liver	disease
lsd	symptoms	lsd	flowering	kidney	increases	kidney	care	kidney	heterosexual	kidney	infections
w_k	$\mathbf{S}_{k,15}^2 \mathbf{S}_{k,118}^2$	w_k	$\mathbf{S}_{k,15}^2 \mathbf{S}_{k,44}^2$	w_k	$\mathbf{S}_{k,16}^2 \mathbf{S}_{k,248}^2$	w_k	$\mathbf{S}_{k,16}^2 \mathbf{S}_{k,147}^2$	w_k	$\mathbf{S}_{k,16}^2 \mathbf{S}_{k,72}^2$	w_k	$\mathbf{S}_{k,16}^2 \mathbf{S}_{k,52}^2$
adhd	3505.8	peyote	3197.5	esophagus	2274.6	aortic	1726.9	erectile	2001.5	abscess	1932.6
anticonvulsants	3047.9	purpurea	2926.6	lobes	1462.4	brainstem	1537.1	dildo	1788.8	multifocal	1440.2
sertraline	2374.6	meo	2878.6	cava	1360.8	endoscopy	1533.7	clitoral	1481.5	hemorrhagic	1239.3
lorazepam	1604.8	deliriants	2397.0	ligamentum	1024.8	laparoscopic	1509.4	deferens	1432.6	esophagitis	1187.4
anticonvulsant	1487.1	diplopterys	2395.1	transversal	1011.2	angioplasty	1447.4	rectal	1352.0	efferent	1160.5
somnolence	1426.1	cabrerana	2237.7	vena	1005.6	cardiology	1433.3	urogenital	1344.7	mitral	1143.5
$\mathrm{E}(S_{16}^2 S_{118}^2) = 2.124$		$\mathrm{E}(S_{30}^2 S_{112}^2) = 2.060$		$\mathrm{E}(S_{30}^2 S_{123}^2) = 1.570$		$\mathrm{E}(S_{36}^2 S_{50}^2) = 1.688$		$\mathrm{E}(S_{36}^2 S_{73}^2) = 1.419$		$\mathrm{E}(S_{168}^2 S_{193}^2) = 1.495$	
Axis 16	Axis 118	Axis 30	Axis 112	Axis 30	Axis 123	Axis 36	Axis 50	Axis 36	Axis 73	Axis 168	Axis 193
blood	disorder	stations	episode	stations	newspaper	http	site	http	ip	license	rights
organs	mental	fm	aired	fm	daily	www	website	www	tcp	copyleft	legislation
liver	disorders	radio	show	radio	weekly	htm	forum	htm	protocols	gpl	act
kidney	symptoms	broadcast	tv	broadcast	newspapers	html	photos	html	protocol	licenses	laws
w_k	$\mathbf{S}_{k,16}^2 \mathbf{S}_{k,118}^2$	w_k	$\mathbf{S}_{k,30}^2 \mathbf{S}_{k,112}^2$	w_k	$\mathbf{S}_{k,30}^2 \mathbf{S}_{k,123}^2$	w_k	$\mathbf{S}_{k,36}^2 \mathbf{S}_{k,50}^2$	w_k	$\mathbf{S}_{k,36}^2 \mathbf{S}_{k,73}^2$	w_k	$\mathbf{S}_{k,168}^2 \mathbf{S}_{k,193}^2$
atrophy	2110.2	rebroadcast	1729.4	canwest	1941.7	shtml	1537.9	mtu	1425.4	copyleft	1648.0
hemiparesis	1877.5	fsn	1635.4	ctv	1343.4	geocities	1230.3	stateful	1352.4	magnatune	1131.9
axonal	1465.9	etv	1600.9	wqxr	1276.0	jeancocteau	871.2	proxying	1324.2	rightsholder	839.7
dysfunction	1380.2	upn	1534.6	superstation	1144.2	lfc	758.0	mpls	1264.5	redistribute	809.7
neuropathy	1300.1	wxyz	1441.3	wanbao	1116.1	uchicago	644.7	vpns	798.6	copyrights	676.5
myopathy	1288.3	whdh	1392.0	aor	998.0	artchive	593.3	kleinrock	796.1	circumvention	653.3
$\mathrm{E}(S_{169}^2 S_{72}^2) = 1.701$		$\mathrm{E}(S_{44}^2 S_{121}^2) = 2.096$		$\mathrm{E}(S_{45}^2 S_{103}^2) = 1.765$		$\mathrm{E}(S_{49}^2 S_{193}^2) = 1.616$		$\mathrm{E}(S_{56}^2 S_{101}^2) = 1.693$		$\mathrm{E}(S_{56}^2 S_{71}^2) = 1.592$	
Axis 169	Axis 72	Axis 44	Axis 121	Axis 45	Axis 103	Axis 49	Axis 193	Axis 56	Axis 101	Axis 56	Axis 71
female	sexual	plants	families	quantum	wavelength	court	rights	cpu	voltage	cpu	game
male	sex	plant	family	particles	light	judge	legislation	microprocessor	electrical	microprocessor	games
age	homosexual	flowers	older	particle	wavelengths	courts	act	processor	circuits	processor	gameplay
infant	heterosexual	flowering	household	physics	laser	trial	laws	cpus	current	cpus	multiplayer
w_k	$\mathbf{S}_{k,169}^2 \mathbf{S}_{k,72}^2$	w_k	$\mathbf{S}_{k,44}^2 \mathbf{S}_{k,121}^2$	w_k	$\mathbf{S}_{k,45}^2 \mathbf{S}_{k,103}^2$	w_k	$\mathbf{S}_{k,49}^2 \mathbf{S}_{k,193}^2$	w_k	$\mathbf{S}_{k,56}^2 \mathbf{S}_{k,101}^2$	w_k	$\mathbf{S}_{k,56}^2 \mathbf{S}_{k,71}^2$
male	1240.5	rosid	4158.0	mesons	3078.7	habeas	954.7	lsi	2589.8	backlit	1398.9
vulval	1191.4	caryophyllales	4064.1	gluons	2079.4	declaratory	823.5	microelectronic	2081.5	quadra	1388.5
faggot	981.2	dicotyledons	3652.5	photon	1648.3	conservatorship	783.1	sram	1901.2	gba	1349.0
spermatozoon	961.5	betulaceae	3637.6	photons	1567.6	waives	753.2	mosfet	1757.0	vectrex	1128.1
frot	940.6	cronquist	3636.1	synchrotron	1435.1	laches	734.4	voltages	1582.9	epyx	1113.6
tribadism	843.5	poaceae	3295.0	isospin	1414.1	talionis	721.3	microcontrollers	1554.5	vcs	1097.0
$\mathrm{E}(S_{188}^2 S_{119}^2) = 1.442$		$\mathrm{E}(S_{188}^2 S_{205}^2) = 1.306$		$\mathrm{E}(S_{67}^2 S_{111}^2) = 1.723$		$\mathrm{E}(S_{67}^2 S_{103}^2) = 1.705$		$\mathrm{E}(S_{101}^2 S_{103}^2) = 1.800$		$\mathrm{E}(S_{101}^2 S_{119}^2) = 1.641$	
Axis 188	Axis 119	Axis 188	Axis 205	Axis 67	Axis 111	Axis 67	Axis 103	Axis 101	Axis 103	Axis 101	Axis 119
less	combustion	less	names	blue	worn	blue	wavelength	voltage	wavelength	voltage	combustion
than	diesel	than	various	white	wearing	white	light	electrical	light	electrical	diesel
more	turbine	more	include	red	clothing	red	wavelengths	circuits	wavelengths	circuits	turbine
much	engine	much	these	yellow	wear	yellow	laser	current	laser	current	engine
w_k	$\mathbf{S}_{k,188}^2 \mathbf{S}_{k,119}^2$	w_k	$\mathbf{S}_{k,188}^2 \mathbf{S}_{k,205}^2$	w_k	$\mathbf{S}_{k,67}^2 \mathbf{S}_{k,111}^2$	w_k	$\mathbf{S}_{k,67}^2 \mathbf{S}_{k,103}^2$	w_k	$\mathbf{S}_{k,101}^2 \mathbf{S}_{k,103}^2$	w_k	$\mathbf{S}_{k,101}^2 \mathbf{S}_{k,119}^2$
thermojet	816.0	gec	505.9	sash	2800.1	thz	2245.4	photodiode	2784.6	alternator	1508.5
dirtier	805.6	nervosa	481.3	hakama	2144.4	monochromatic	1777.2	diodes	2320.6	cogeneration	1485.3
cng	739.9	dddddd	430.1	turbans	1756.4	protanomaly	1545.8	diode	1959.3	clamp	1370.2
oxidiser	655.3	ler	373.6	waists	1745.6	fluoresce	1376.9	ssi	1633.5	solenoid	1183.5
tuyere	634.7	gomoku	356.8	tunics	1612.8	scintillator	1366.4	bandgap	1523.4	thermionic	1144.8
propfan	630.0	kel	317.6	pleated	1561.4	diffuser	1233.8	photodiodes	1478.2	transformer	1055.4

Table 8: Complementary experimental results to Table 3. For all component pairs (S_i, S_j) in the first subtree of the MST in Fig. 5, the top 6 words and their corresponding $\mathbf{S}_{t,i}^2 \mathbf{S}_{t,j}^2$ values that contribute the most to the $\mathrm{E}(S_i^2 S_j^2)$ value are presented.

$\mathrm{E}(S_2^2 S_{10}^2) = 2.323$		$\mathrm{E}(S_2^2 S_{119}^2) = 1.755$		$\mathrm{E}(S_{132}^2 S_{30}^2) = 1.820$		$\mathrm{E}(S_{132}^2 S_{73}^2) = 1.693$		$\mathrm{E}(S_{132}^2 S_{62}^2) = 1.632$		$\mathrm{E}(S_{136}^2 S_{56}^2) = 1.967$	
Axis 2	Axis 10	Axis 2	Axis 119	Axis 132	Axis 30	Axis 132	Axis 73	Axis 132	Axis 62	Axis 136	Axis 56
acid	dna	acid	combustion	telephone	stations	telephone	ip	telephone	company	disk	cpu
hydrogen	proteins	hydrogen	diesel	phone	fm	phone	tcp	phone	corporation	floppy	microprocessor
acids	rna	acids	turbine	mobile	radio	mobile	protocols	mobile	companies	disks	processor
oh	mrna	oh	engine	cellular	broadcast	cellular	protocol	cellular	shareholders	drives	cpus
w_k	$\mathbf{S}_{k,2}^2 \mathbf{S}_{k,10}^2$	w_k	$\mathbf{S}_{k,2}^2 \mathbf{S}_{k,119}^2$	w_k	$\mathbf{S}_{k,132}^2 \mathbf{S}_{k,30}^2$	w_k	$\mathbf{S}_{k,132}^2 \mathbf{S}_{k,73}^2$	w_k	$\mathbf{S}_{k,132}^2 \mathbf{S}_{k,62}^2$	w_k	$\mathbf{S}_{k,136}^2 \mathbf{S}_{k,56}^2$
ribose	3755.7	pyrolysis	2794.4	digitalized	4726.8	multipoint	2062.7	esat	3145.5	sata	3427.5
deoxyribose	2963.9	syngas	2056.9	arabsat	4657.2	pstn	1996.9	telecoms	2547.9	udma	2519.7
phosphodiester	2850.2	gasification	1783.5	radiotelephone	4453.9	wimax	1873.8	nynex	2155.9	backplanes	2008.9
biosynthesis	2510.1	butane	1761.2	landlines	3522.1	xdsl	1491.2	gnc	1810.0	nexgen	1947.7
methyltransferase	2482.9	dehydrogenation	1623.0	intersputnik	3053.5	svcs	1361.2	haitel	1657.3	megabytes	1890.1
pyrimidine	2399.6	tert	1230.2	telex	2722.4	isdn	1235.8	openreach	1529.2	eisa	1859.5

$\mathrm{E}(S_{10}^2 S_{16}^2) = 1.947$		$\mathrm{E}(S_{10}^2 S_{160}^2) = 1.811$		$\mathrm{E}(S_{140}^2 S_{56}^2) = 1.615$		$\mathrm{E}(S_{13}^2 S_{56}^2) = 1.732$		$\mathrm{E}(S_{13}^2 S_{168}^2) = 1.991$		$\mathrm{E}(S_{13}^2 S_{73}^2) = 1.740$	
Axis 10	Axis 16	Axis 10	Axis 160	Axis 140	Axis 56	Axis 13	Axis 56	Axis 13	Axis 168	Axis 13	Axis 73
dna	blood	dna	evolution	import	cpu	windows	cpu	windows	license	windows	ip
proteins	organs	proteins	evolutionary	duplicate	microprocessor	os	microprocessor	os	copyleft	os	tcp
rna	liver	rna	darwin	info	processor	unix	processor	unix	gpl	unix	protocols
mrna	kidney	mrna	selection	no	cpus	linux	cpus	linux	licenses	linux	protocol
w_k	$\mathbf{S}_{k,10}^2 \mathbf{S}_{k,16}^2$	w_k	$\mathbf{S}_{k,10}^2 \mathbf{S}_{k,160}^2$	w_k	$\mathbf{S}_{k,140}^2 \mathbf{S}_{k,56}^2$	w_k	$\mathbf{S}_{k,13}^2 \mathbf{S}_{k,56}^2$	w_k	$\mathbf{S}_{k,13}^2 \mathbf{S}_{k,168}^2$	w_k	$\mathbf{S}_{k,13}^2 \mathbf{S}_{k,73}^2$
adenylate	2079.8	utr	2381.5	superpipelined	5652.9	xcode	2609.7	qpl	5678.2	netware	1799.2
effectors	1842.5	reticulum	1942.0	strongarm	3220.2	powerpc	2046.7	lgpl	4519.9	netbios	1543.2
antisense	1639.9	genomic	1668.6	specrate	2470.5	itanium	1500.1	trolltech	3588.4	imap	1414.0
cyclase	1638.9	homozygous	1599.1	specbaserate	1524.5	glibc	1355.1	gpl	3325.2	glut	1239.0
myosin	1201.8	cleaved	1181.0	insubstantial	1387.8	irix	1177.1	gnu	2826.1	wfw	1179.9
axons	1144.2	tubulin	1152.4	eisa	1027.8	efi	1161.1	bsd	2822.7	dhcpv	1115.5

$\mathrm{E}(S_{15}^2 S_{118}^2) = 1.959$		$\mathrm{E}(S_{15}^2 S_{44}^2) = 1.793$		$\mathrm{E}(S_{16}^2 S_{248}^2) = 1.499$		$\mathrm{E}(S_{16}^2 S_{147}^2) = 1.717$		$\mathrm{E}(S_{16}^2 S_{72}^2) = 1.628$		$\mathrm{E}(S_{16}^2 S_{52}^2) = 2.158$	
Axis 15	Axis 118	Axis 15	Axis 44	Axis 16	Axis 248	Axis 16	Axis 147	Axis 16	Axis 72	Axis 16	Axis 52
drugs	disorder	drugs	plants	blood	increase	blood	medicine	blood	sexual	blood	infectious
drug	mental	drug	plant	organs	increased	organs	medical	organs	sex	organs	infection
heroin	disorders	heroin	flowers	liver	increasing	liver	doctors	liver	homosexual	liver	disease
lsd	symptoms	lsd	flowering	kidney	increases	kidney	care	kidney	heterosexual	kidney	infections
w_k	$\mathbf{S}_{k,15}^2 \mathbf{S}_{k,118}^2$	w_k	$\mathbf{S}_{k,15}^2 \mathbf{S}_{k,44}^2$	w_k	$\mathbf{S}_{k,16}^2 \mathbf{S}_{k,248}^2$	w_k	$\mathbf{S}_{k,16}^2 \mathbf{S}_{k,147}^2$	w_k	$\mathbf{S}_{k,16}^2 \mathbf{S}_{k,72}^2$	w_k	$\mathbf{S}_{k,16}^2 \mathbf{S}_{k,52}^2$
adhd	3505.8	peyote	3197.5	esophagus	2274.6	aortic	1726.9	erectile	2001.5	abscess	1932.6
anticonvulsants	3047.9	purpurea	2926.6	lobes	1462.4	brainstem	1537.1	dildo	1788.8	multifocal	1440.2
sertraline	2374.6	meo	2878.6	cava	1360.8	endoscopy	1533.7	clitoral	1481.5	hemorrhagic	1239.3
lorazepam	1604.8	deliriants	2397.0	ligamentum	1024.8	laparoscopic	1509.4	deferens	1432.6	esophagitis	1187.4
anticonvulsant	1487.1	diplopterys	2395.1	transversal	1011.2	angioplasty	1447.4	rectal	1352.0	efferent	1160.5
somnolence	1426.1	cabrerana	2237.7	vena	1005.6	cardiology	1433.3	urogenital	1344.7	mitral	1143.5

$\mathrm{E}(S_{16}^2 S_{118}^2) = 2.124$		$\mathrm{E}(S_{30}^2 S_{112}^2) = 2.060$		$\mathrm{E}(S_{30}^2 S_{123}^2) = 1.570$		$\mathrm{E}(S_{36}^2 S_{50}^2) = 1.688$		$\mathrm{E}(S_{36}^2 S_{73}^2) = 1.419$		$\mathrm{E}(S_{168}^2 S_{193}^2) = 1.495$	
Axis 16	Axis 118	Axis 30	Axis 112	Axis 30	Axis 123	Axis 36	Axis 50	Axis 36	Axis 73	Axis 168	Axis 193
blood	disorder	stations	episode	stations	newspaper	http	site	http	ip	license	rights
organs	mental	fm	aired	fm	daily	www	website	www	tcp	copyleft	legislation
liver	disorders	radio	show	radio	weekly	htm	forum	htm	protocols	gpl	act
kidney	symptoms	broadcast	tv	broadcast	newspapers	html	photos	html	protocol	licenses	laws
w_k	$\mathbf{S}_{k,16}^2 \mathbf{S}_{k,118}^2$	w_k	$\mathbf{S}_{k,30}^2 \mathbf{S}_{k,112}^2$	w_k	$\mathbf{S}_{k,30}^2 \mathbf{S}_{k,123}^2$	w_k	$\mathbf{S}_{k,36}^2 \mathbf{S}_{k,50}^2$	w_k	$\mathbf{S}_{k,36}^2 \mathbf{S}_{k,73}^2$	w_k	$\mathbf{S}_{k,168}^2 \mathbf{S}_{k,193}^2$
atrophy	2110.2	rebroadcast	1729.4	canwest	1941.7	shtml	1537.9	mtu	1425.4	copyleft	1648.0
hemiparesis	1877.5	fsn	1635.4	ctv	1343.4	geocities	1230.3	stateful	1352.4	magnatune	1131.9
axonal	1465.9	etv	1600.9	wqxr	1276.0	jeancocteau	871.2	proxying	1324.2	rightsholder	839.7
dysfunction	1380.2	upn	1534.6	superstation	1144.2	lfc	758.0	mpls	1264.5	redistribute	809.7
neuropathy	1300.1	wxyz	1441.3	wanbao	1116.1	uchicago	644.7	vpns	798.6	copyrights	676.5
myopathy	1288.3	whdh	1392.0	aor	998.0	artchive	593.3	kleinrock	796.1	circumvention	653.3

$\mathrm{E}(S_{169}^2 S_{72}^2) = 1.701$		$\mathrm{E}(S_{44}^2 S_{121}^2) = 2.096$		$\mathrm{E}(S_{45}^2 S_{103}^2) = 1.765$		$\mathrm{E}(S_{49}^2 S_{193}^2) = 1.616$		$\mathrm{E}(S_{56}^2 S_{101}^2) = 1.693$	
Axis 169	Axis 72	Axis 44	Axis 121	Axis 45	Axis 103	Axis 49	Axis 193	Axis 56	Axis 101
female	sexual	plants	families	quantum	wavelength	court	rights	cpu	voltage
male	sex	plant	family	particles	light	judge	legislation	microprocessor	electrical
age	homosexual	flowers	older	particle	wavelengths	courts	act	processor	circuits
infant	heterosexual	flowering	household	physics	laser	trial	laws	cpus	current
w_k	$\mathbf{S}_{k,169}^2 \mathbf{S}_{k,72}^2$	w_k	$\mathbf{S}_{k,44}^2 \mathbf{S}_{k,121}^2$	w_k	$\mathbf{S}_{k,45}^2 \mathbf{S}_{k,103}^2$	w_k	$\mathbf{S}_{k,49}^2 \mathbf{S}_{k,193}^2$	w_k	$\mathbf{S}_{k,56}^2 \mathbf{S}_{k,101}^2$
male	1240.5	rosid	4158.0	mesons	3078.7	habeas	954.7	lsi	2589.8
vulval	1191.4	caryophyllales	4064.1	gluons	2079.4	declaratory	823.5	microelectronic	2081.5
faggot	981.2	dicotyledons	3652.5	photon	1648.3	conservatorship	783.1	sram	1901.2
spermatozoon	961.5	betulaceae	3637.6	photons	1567.6	waives	753.2	mosfet	1757.0
frot	940.6	cronquist	3636.1	synchrotron	1435.1	laches	734.4	voltages	1582.9
tribadism	843.5	poaceae	3295.0	isospin	1414.1	talionis	721.3	microcontrollers	1554.5

Table 9: Complementary experimental results to Table 3. For all component pairs (S_i, S_j) in the second subtree of the MST in Fig. 5, the top 6 words and their corresponding $\mathbf{S}_{t,i}^2 \mathbf{S}_{t,j}^2$ values that contribute the most to the $\mathrm{E}(S_i^2 S_j^2)$ value are presented.