From the New World of Word Embeddings: A Comparative Study of Small-World Lexico-Semantic Networks in LLMs

Anonymous ACL submission

Abstract

Lexico-semantic networks represent words as nodes and their semantic relatedness as edges. While such networks are traditionally constructed using embeddings from encoder-based models or static vectors, embeddings from decoder-only large language models (LLMs) remain underexplored. Unlike encoder models, LLMs are trained with a next-token prediction objective, which does not directly encode the meaning of the current token. In this paper, we construct lexico-semantic networks from the input embeddings of LLMs with varying parameter scales and conduct a comparative analysis of their global and local structures. Our results show that these networks exhibit small-world properties, characterized by high clustering and short path lengths. Moreover, larger LLMs yield more intricate networks with less small-world effects and longer paths, reflecting richer semantic structures and relations. We further validate our approach through analyses of common conceptual pairs, structured lexical relations derived from WordNet, and a cross-lingual semantic network for qualitative words.

1 Introduction

What does the meaning of a word come from? According to the distributional hypothesis (Harris, 1954; Firth, 1957; Boleda, 2020), it arises from the relationships of the word with other words. These relationships - whether syntagmatic (co-occurrence within the same context) or paradigmatic (sharing similar neighboring words) - can be modeled as a network connecting different but related words. This distributional perspective on meaning, along with network-based representations, has been widely explored in areas such as semantic lexicons (e.g., WordNet-inspired resources) (Miller, 1995), language typology (e.g., semantic map models) (Haspelmath, 2003) and cognitive science (e.g., conceptual space framework) (Gärdenfors, 2000).

Lexico-semantic networks can be built manually based on conceptual spaces (Gärdenfors, 2000) or induced from cross-linguistic correspondences (Croft, 2001), but both face scalability challenges. Word embeddings offer a scalable alternative, capturing semantic relations like analogy and similarity (Mikolov et al., 2013; Vulić et al., 2020). However, prior work often relies on static or encoder-based models (e.g., word2vec (Mikolov et al., 2013), BERT (Devlin et al., 2019)), overlooking decoder-only LLMs and the global structure encoded in their full vocabulary space.

These considerations lead to an intriguing question: what do the semantic networks induced from the *new LLM world* look like? It is "new" in the sense that input embeddings from LLMs remain relatively underexplored, as their next-token prediction objective does not directly encode the meaning of the current token (Liu et al., 2024). It is a "world" because we aim to construct networks over the *entire vocabulary*, offering a global perspective on the internal structure of LLMs. Furthermore, how do such networks differ across LLMs that share the same architecture but vary in computational scale? Given the general trend, established by the scaling law (Kaplan et al., 2020), that larger models tend to perform better, we ask: is this also reflected in the structure of their induced lexico-semantic networks?

We construct a lexico-semantic network from input embeddings of large language models (LLMs), treating vocabulary embeddings as nodes and defining edge weights via similarity metrics. Starting from a fully connected graph, we prune weak edges based on the connectivity hypothesis from semantic map models (Haspelmath, 2003; Liu et al., 2025), preserving overall connectivity. We analyze the resulting networks across LLMs of different scales, finding small-world structures characterized by high clustering and short path lengths, with larger models exhibiting denser and more complex patterns. To assess practical utility, we examine local subgraphs in three scenarios—common conceptual

domains, WordNet-based lexical relations, and a cross-lingual case study—showing strong alignment with human-annotated resources and validating the effectiveness of our approach.

In conclusion, our main contributions are as follows:

- We construct a lexico-semantic network using the input embeddings of the entire vocabulary from large language models (LLMs), and show that these networks exhibit the small-world property, indicating strong local clustering and efficient word-to-word connectivity.
- We conduct a comparative analysis of lexico-semantic networks derived from LLMs of different parameter scales, grounded in the connectivity hypothesis. Our results reveal that larger-scale models display a weaker small-world effect, characterized by longer average path lengths and more complex community structures.
- We design three evaluation scenarios to analyze the lexico-semantic networks from both global and local perspectives: (1) common conceptual domains, (2) WordNet-based lexical relations, and (3) a cross-lingual case study on evaluative words. In all cases, larger-scale LLMs consistently yield networks with longer semantic paths, reflecting greater relational richness and more nuanced concept structures.

2 Related Work

2.1 Word Embedding and Representation

Contemporary language models represent words or subtokens using continuous vectors based on distributed semantics (Boleda, 2020). These high-dimensional vectors can be static (Mikolov et al., 2013; Bojanowski et al., 2017) or context-sensitive (Devlin et al., 2019). While effective, they lack the interpretability of semantic feature-based representations (Petersen and Potts, 2023). Word embeddings exhibit elegant linear relationships (Mikolov et al., 2013), high similarity with human judgments (Vulić et al., 2020), and meaningful representations (Turney and Pantel, 2010). Static embeddings are particularly suitable for offline, context-free words, especially monosemous ones. These embeddings are used in the input and output layers of LLMs. Previous research has explored their linear properties (Han et al., 2024), conceptual space construction (Moullec and Douven, 2025), similarity distributions on the tasks of word-pair similarity and analogy (Freestone and Santu, 2024) and other aspects. They focus on several instances rather than the network structure built by the whole vocabulary. In this paper, we explore the lexico-semantic network via LLM input embeddings, offering a more global and systematic evaluation.

2.2 Lexico-semantic Network Representations

A prominent meaning-driven approach to constructing lexico-semantic networks is the *conceptual space framework* (CSF), widely applied in cognitive science (Gärdenfors, 2000, 2014; Nosofsky, 1986) and neuroscience (Caglar, 2021). CSF models concepts as points or regions in a continuous multidimensional space. For example, the color domain is structured by hue, saturation, and brightness, with color terms corresponding to convex regions (Gärdenfors, 2000). Similar structures have been proposed for other domains, such as body parts and kinship (Zwarts, 2010).

A data-driven alternative is *semantic map modeling* (SMM), which builds networks from cross-linguistic co-expression patterns found in content words (Guo, 2012; Petersen and Potts, 2023; Dellert, 2024), function words (Zhang, 2017), or constructions (Malchukov et al., 2007). Nodes typically represent grammatical or conceptual meanings.

Networks can be built via bottom-up or top-down strategies. The bottom-up method relies on the *connectivity hypothesis* (Haspelmath, 2003), linking concepts co-expressed by a single form. The top-down approach (Liu et al., 2025) begins with a complete graph weighted by semantic similarity, then prunes it to retain informative edges.

In this paper, we adopt a data-driven, top-down approach to lexico-semantic network construction, using LLM input embeddings and treating embedding similarity as a proxy for semantic relatedness.

3 Approach

In this section, we first introduce the basic notions and construct a complete graph connecting all conceptual nodes. We then sparsify the graph based on the revised connectivity hypothesis. Finally, we define global and local metrics to evaluate the lexico-semantic network. The overall pipeline is illustrated in Figure 1.

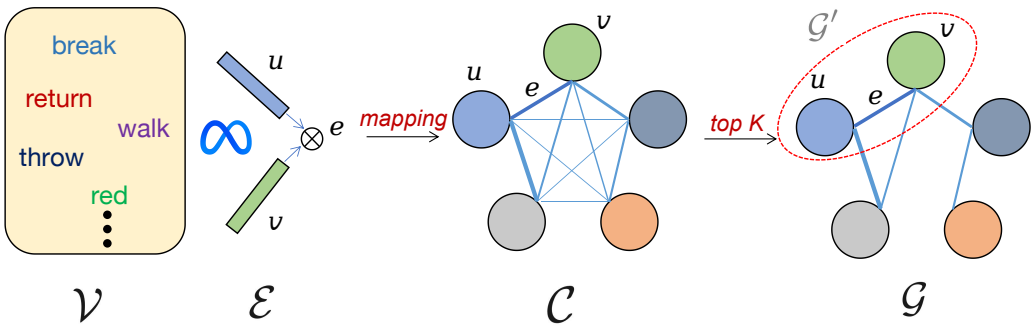


Figure 1: Outline of our lexico-semantic network construction. First, we extract the input word embeddings ($\mathcal{E}$) for the LLM vocabulary ($\mathcal{V}$). Next, we build a complete graph $\mathcal{C}$ by calculating the cosine similarity between all embedding pairs. Finally, we retain edges based on similarity, from highest to lowest, until the graph $\mathcal{G}$ is connected. We then focus on specific connected subgraphs $\mathcal{G}'$ representing certain domains at a local level.

3.1 Basic Notions

We define a lexico-semantic network $\mathcal{G} = \{V, E\}$, where V and E are sets of nodes and edges, respectively. Each node $v \in V$ represents a word or a token [1]. Each edge $e(u, v) \in E$ connects a pair of nodes (u, v), reflecting their semantic relatedness. If a path $p(u, v)$ exists between nodes u and v, they are connected, with path length L defined as the number of edges along the path. If no path exists, $L = \infty$. We use L of the shortest path to indicate the topological distance between two nodes. The network $\mathcal{G}$ is considered connected if every pair of nodes is connected.

A subgraph $\mathcal{G}' = \{V', E'\}$, where $V' \subset V$ and $E' \subset E$, reflects the local topology of $\mathcal{G}$ and typically represents a specific semantic domain, such as adverbs (Zhang, 2017), color adjectives (Gärdenfors, 2014), or qualitative words (Perrin, 2010). Similarly, a subgraph is connected if every pair of nodes has a path. Note that $\mathcal{G}'$ may not connected as a whole.

We define a metric M on $\mathcal{G}$ to measure the relatedness or similarity between nodes. A common metric is cosine similarity , widely applied in similarity-related tasks. Accordingly, we also utilize the cosine distance D between nodes, which is one minus cosine similarity [2].

3.2 Complete Graph

We use an LLM to extract input embeddings $\mathcal{E}$ for all tokens in its vocabulary $\mathcal{V}$, treating each token (the minimal computational unit) as a node in the lexico-semantic network. After obtaining the vectorized embeddings, we compute the cosine distance between every pair of nodes to define edge weights. Additionally, we apply centering by subtracting the average vector from each embedding to address anisotropy (**?**). This results in a complete graph $\mathcal{C}$, where every pair of words is connected.

3.3 Lexico-semantic Space

We derive a sparsified graph, denoted as the lexico-semantic space $\mathcal{G}$, from the complete graph $\mathcal{C}$. We propose a minimum connectivity hypothesis, which states that $\mathcal{G}$ must remain connected while using the fewest edges possible. The "connectivity" ensures that every pair of words is connected, forming a valid space. The "minimum" condition favors sparse connections, inspired by the top-down construction of semantic map models, which even use trees (with the least number of edges) to maintain connectivity (Liu et al., 2025). To achieve this sparsity, we rank edges by weight defined on the metric M and retain the top K ratio of edges, as higher weights indicate more important connections.

A well-defined $\mathcal{G}$ is also a discrete topological space $\{\mathcal{G}, \mathcal{T}\}$, where $\mathcal{T}$ is the collection of all subsets. We define a subgraph $\mathcal{G}'$ as a subset of $\mathcal{G}$ and it is considered an open set. This is because the intersection and union of any two subgraphs $\mathcal{G}_\mathcal{A}$ and $\mathcal{G}_\mathcal{B}$ still belong to $\mathcal{T}$:

$$\forall \mathcal{G}_A, \mathcal{G}_B \in \mathcal{T}, \quad \mathcal{G}_A \cap \mathcal{G}_B \in \mathcal{T}, \quad \mathcal{G}_A \cup \mathcal{G}_B \in \mathcal{T}. \quad (1)$$

This is ensured by the "connectivity" condition, while a "minimum" topological space is required for the conceptual structure.

[1] In LLMs, a token, typically a subword, is the minimal computational unit.

[2] While cosine distance violates the triangle inequality required by strict distances, we relax this constraint due to its simplicity and widespread use.

3.4 Evaluation

We evaluate the lexico-semantic network from both global and local perspectives.

Globally, we compute network statistics to analyze basic properties, connectivity, and small-world characteristics of the spaces built by two models. Small-world characteristics are indicated by a higher clustering coefficient and a shorter shortest path, which are described in detail in Section 5.1.

Locally, we analyze a subgraph $\mathcal{G}'$ of the conceptual space $\mathcal{G}$ in three scenarios. Scenario 1 examines common concepts across ten semantic categories, each containing monosemous words, comparing shortest paths within and between groups. Scenario 2 explores shortest-path connections for various WordNet relations. Scenario 3 evaluates a lexico-semantic network of qualitative words, comparing it to the corresponding LLM subgraph. Beyond topology and connectivity, we assess node degree correlations and measure recall and precision against the ground truth.

4 Experimental Design

4.1 Large Language Models

We adopt the Llama series as our LLMs, including Llama2-7B and Llama2-70B (Touvron et al., 2023). The dimension of the input embedding is 4096 and 8192, for Llama2-7B and Llama2-70B respectively. Also, they share the vocabulary for both models, with the size of vocabulary 32,000. The tokens in the vocabulary are obtained by Byte Pair Encoding (Sennrich et al., 2016), merging the bigram with the most frequent co-occurrence. Thus, many tokens are part of a whole word. Besides, tokens at the beginning of a word are different from those in other places, i.e., the end part of a word. For example, "man" in "policeman" and "man" are different units in the vocabulary. We identify the token appearing the end part of a word by add "#" at the beginning of the token, such as "#man".

4.2 Scenario 1: Common Concepts

We construct nine semantic groups representing common concepts, each containing ten frequent words: NUMBER, NAME, MONTH, COLOR, CITY, NATION, PLACE, HUMAN, and FURNITURE. Some of these groups, such as NUMBER, NAME, and MONTH, naturally exhibit sequential structures. Additionally, we include a RANDOM group as a control. All words in these groups are monosemous and consist of a single token, ensuring that they appear directly in the vocabulary and have well-defined type-level meanings. A full list of the concepts is provided in Table 4 in Appendix A.1. This scenario investigates how well the embedding-based network captures conceptual similarity, as reflected by the length of shortest paths within and across semantic groups.

4.3 Scenario 2: WordNet Relations

In Scenario 2, we examine a subgraph of WordNet (Miller, 1995) to assess how structural lexical relations are reflected in the embedding-based network. We use a filtered subset of the public WN18 dataset (Bordes et al., 2013), comprising 40,943 WordNet synsets and 18 relation types. The filtering process follows these criteria: (1) after mapping synsets to their corresponding words, only those present in the LLaMA vocabulary are retained; (2) only the first sense of each synset is kept to represent its prototypical meaning; (3) only relation types involving at least 10 valid word pairs are included; (4) symmetric relation types, such as *hypernym* and *hyponym*, are merged.

We also introduce two additional wordform-based relation types: *tokenization variant*, which captures differences between tokens with and without a leading underscore (e.g., "man" vs. "#man"), and *uppercase variant*, which differentiates capitalized from lowercase forms (e.g., "red" vs. "Red"). In total, we include eight relation types, summarized in Table 1. This scenario investigates whether and how these structured lexical relations are preserved in the topology of the embedding-based semantic network.

Index	Relation Type	Count
A	Member of Domain Topic	51
B	Verb Group	10
C	Hypernym	464
D	Has Part	22
E	Also See	95
F	Derivationally Related Form	388
G	Tokenization Variant	1685
H	Uppercase Variant	1788

Table 1: WordNet Relations and Instance Counts

4.4 Scenario 3: SMM of Qualitative Words

In Scenario 3, we leverage a cross-linguistic semantic map of adjectives and qualitative terms (Perrin,

2010), which encompasses 22 African languages along with French and English. In this resource, polysemous words within each language are linked to reflect conceptual proximity, forming a semantic map that encodes a universal network of meanings. This map captures cross-linguistic regularities in polysemy patterns, serving as an invariant conceptual structure underlying diverse lexical realizations. The semantic domains covered include dimension, age, value, and color, with concepts represented by capitalized English words (e.g., *BIG*, *SMALL*, *LONG*, *SHORT*, *WIDE*, *DEEP* for the dimension domain).

We filter out words not present in the LLaMA vocabulary, resulting in a final set of 75 concepts from the original 110. The full list of concepts and the human-annotated semantic graph are provided in Appendix A.2. This scenario evaluates how well the embedding-induced lexico-semantic network aligns with a cross-linguistically grounded conceptual structure.

5 Results and Analysis

In this section, we first construct the lexico-semantic network based on the minimum connectivity approach described in Section 3.3. We then evaluate the space in three distinct scenarios.

5.1 Graph Construction

Choice of K. To ensure the graph is minimally connected, we extract the top K ratio of edges, where the edge weights are determined by cosine similarity. We incrementally increase the value of K while monitoring the number of connected components (CC), as shown in Figure 2. The graph first becomes connected when the log of the number of CC reaches zero. In our experiments, we selected $K = 0.002$, at which point both models become connected. This choice ensures that the same number of edges are used for both models, making the comparison fair.

Global Statistics. We present the statistics of the lexico-semantic networks for both models in Table 2, divided into three parts. The **Basic** section includes the number of nodes (#Nodes), edges (#Edges), and the average (Avg. Degree) and standard deviation (Std. Degree) of degrees. The **Weighted** section calculates the average and standard deviation of weighted degrees (also called Strength or Traffic), along with the equivalent threshold—the minimum weight value in the final graph G. The last section focuses on **Small-world** effects. The global clustering coefficient (GCC) measures graph transitivity—the fraction of actual triangles among all possible triangles in G. The local clustering coefficient (ALCC) is calculated as the average of actual connections within neighbors for all nodes. The network diameter is the longest path between any two nodes. Smaller diameters and larger GCC/ALCC values indicate stronger

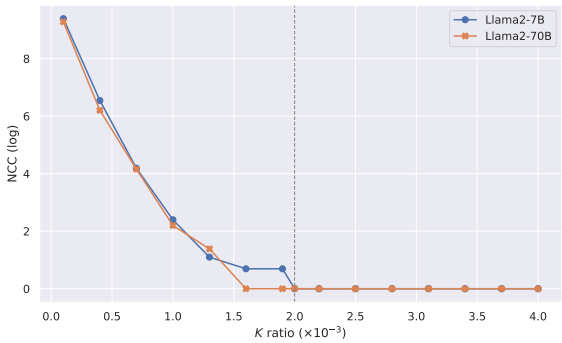


Figure 2: Logarithm of the number of connected components as the top K ratio increases for Llama2-7B and Llama2-70B. A value of zero indicates a fully connected network, while the dotted line marks the first ratio at which both models become connected.

Statistics	Llama2-7B	Llama2-70B
Basic		
#Nodes	32,000	32,000
#Edges	1,024,000	1,024,000
Avg. Degree	64	64
Std. Degree	68.39	58.96
Weighted		
Avg. Degree_W	8.76*	12.23*
Std. Degree_W	13.78	14.02
Threshold	0.095	0.147
Small-world		
GCC (↑)	0.325	0.215
ALCC (↑)	0.183*	0.174*
Diameter (↓)	6	6

Table 2: Statistics for Llama2-7B and Llama2-70B. GCC refers to global clustering coefficient, ALCC to average local clustering coefficient, and Diameter refers to the longest path length. A star (*) indicates high statistical significance between the two models for that metric. Metrics without a star are not tested for significance, typically because they are fixed by design (e.g., #Nodes, #Edges) or not applicable for statistical comparison.

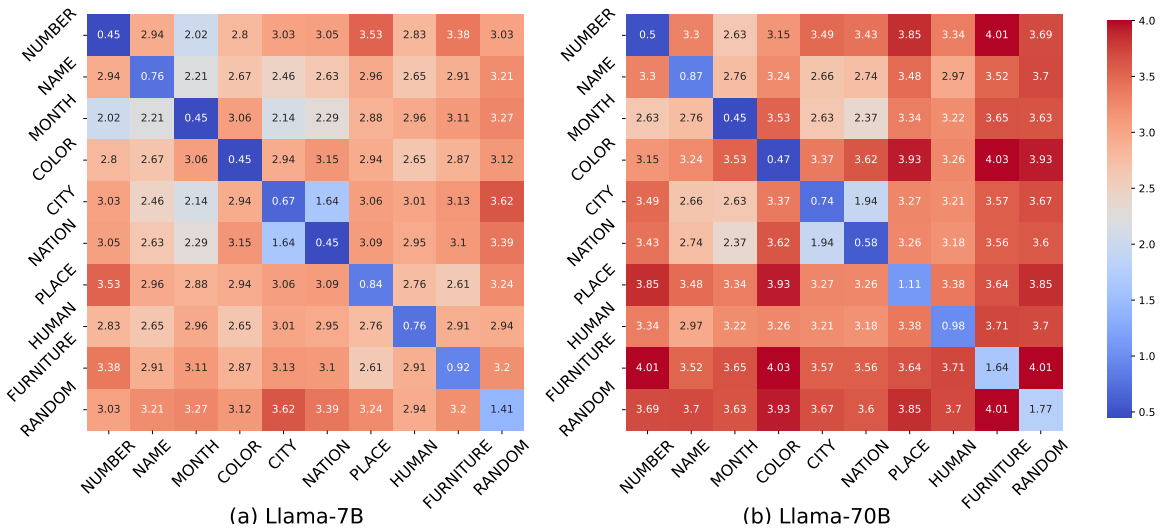


Figure 3: Shortest path lengths among semantic groups for Llama2-7B (left) and Llama2-70B (right).

small-world effects. We also calculate the statistical difference of t-test and mark it by a star (*) for a high difference (with a p value less than 0.05).

Llama2-7B and Llama2-70B share similar graph structures, both using the same number of edges. The average degree is 64, but 7B has a flatter degree distribution with a larger standard deviation. In the weighted version, 7B's degree distribution is more concentrated. As shown in Figure 9 (Appendix A.3), 7B exhibits a long-tailed distribution, indicating fewer high-degree (central) nodes.

Both models exhibit pronounced small-world characteristics, reflected in their high global clustering coefficients (GCC), average local clustering coefficients (ALCC), and short diameters. In contrast, random networks with the same number of edges show substantially lower GCC (0.0032) and ALCC (0.0020). Notably, the observed diameter of 6 corresponds with the Six Degrees of Separation theory (Milgram, 1967), a phenomenon commonly observed in social networks (Watts and Strogatz, 1998) and web graphs (Albert et al., 1999). Compared to the smaller model, the 70B model displays a somewhat weaker small-world effect, evidenced by higher GCC and ALCC values, suggesting longer and more complex node interactions. Our subsequent local analyses provide further support for these observations through concrete examples.

5.2 Scenario 1

In this scenario, we evaluate the lexico-semantic networks of both models by calculating the shortest path length between any pair of nodes within

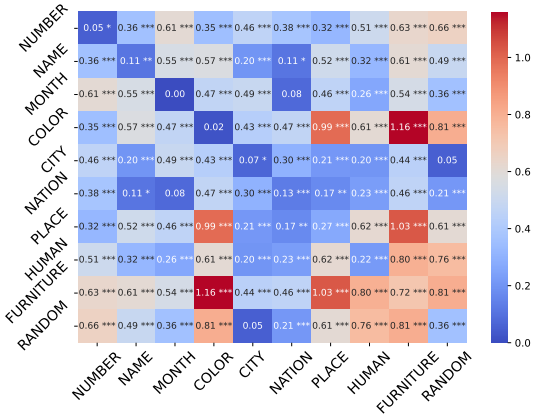


Figure 4: Shortest Path Length Difference (Llama2-70B Minus Llama2-7B) Across Semantic Groups. The number of stars indicate the degree of the significance level of the difference.

and between semantic groups. The shortest path is computed using Dijkstra's algorithm from the NetworkX package [3] with the cosine distance between any two nodes. The average lengths for Llama2-7B and Llama2-70B are shown in Figure 3. Figure 4 presents the difference heatmap, where the difference is computed as the average length of Llama2-70B minus that of Llama2-7B. The significance of the differences is indicated by the number of stars (one, two, or three), corresponding to p-values of 0.05, 0.01, and 0.001 in the t-test, respectively.

Our results indicate that Llama2-70B generally exhibits longer path lengths than Llama2-7B, both within and across semantic groups, suggesting that

[3] https://networkx.org/documentation/stable/reference/algorithms/shortest_paths.html

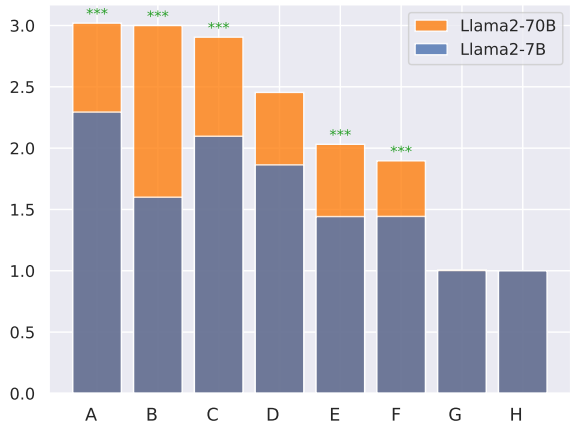


Figure 5: Averaged shortest path length across relation types for both models. The number of stars indicates the significance of the difference.

Llama2-70B uncovers relationships through more complex pathways. Figures 6a and 6b illustrate the six shortest paths between "bar" and "library." The shortest path (highlighted in red) in Llama2-70B shows greater variation in word forms and conjugations. Moreover, alternative paths involve multilingual links and diverse connections that assist in disambiguation—for instance, differentiating the building "library" from its software-related sense. This richer semantic representation may account for the superior performance of larger-scale models.

When analyzing per group, both models show shorter paths within the same group, and the trends are similar for different group pairs. Among the groups, COLOR shows the shortest paths, while FURNITURE exhibits the longest. This may be due to polysemy in FURNITURE, where terms like "chair" refer to both furniture and a human (e.g., a person referred to as a "chair"). This is further supported by the longer paths when comparing FURNITURE to other groups. Conversely, NATION and CITY tend to have shorter paths.

5.3 Scenario 2

In the second scenario, we evaluate the network for word pairs from different relation types, as shown in Table 1. For each pair, we compute the shortest path and display the length for both models. The significance of the differences is indicated by the number of stars. The results are shown in Figure 5.

The results mirror those of Scenario 1: Llama2-70B generally exhibits longer paths than Llama2-7B. For most relation types, the path lengths in 70B are significantly greater, except for word form–related relations where both models maintain direct edges between word pairs. Longer paths are observed in relations such as *member of domain topic*, *verb group*, and *hypernym*, reflecting more complex connection chains. For instance, in the *hypernym* relation between "set" and "collection," Llama2-70B demonstrates a more nuanced path, gradually linking the singular "set" to the plural "collection" with conjugation correlations, as illustrated in Figures 7a and 7b. This suggests that larger-scale models capture more logical and intricate conceptual relations.

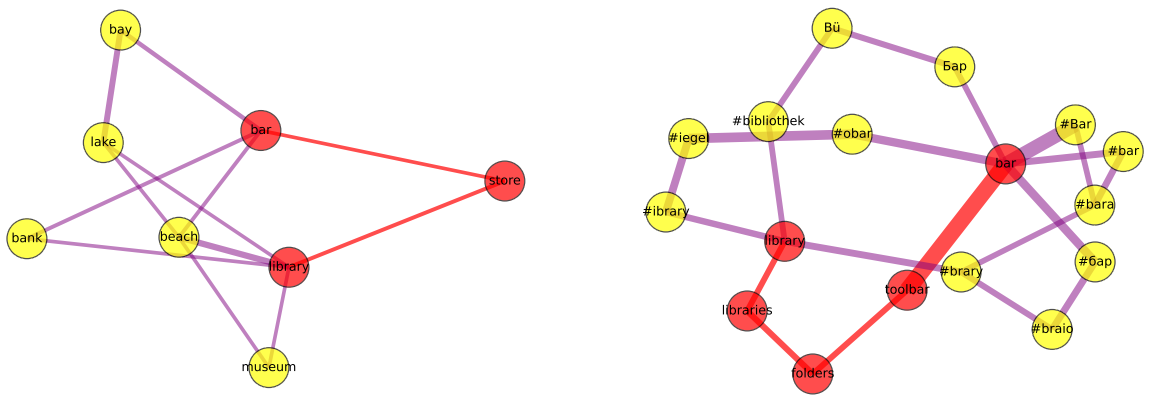
5.4 Scenario 3

In Scenario 3, we construct a lexico-semantic network for qualitative words, with a reference (GT) from cross-lingual research. The statistics for the spaces built by Llama2-7B (7B), Llama2-70B (70B), and GT are shown in Table 3.

Statistics	**7B**	**70B**	**GT**
Basic			
#Nodes	75	75	75
#Edges	293	130	37
Avg. Degree	7.813	3.467	0.987
Std. Degree	5.724	3.021	0.959
Weighted			
Avg. Degree_W	0.963	0.611	-
Std. Degree_W	0.736	0.546	-
Avg. Weight	0.123	0.176	-
Connectivity			
#Component (↓)	8	17	39
#Single (↓)	7	16	26
Reference with GT			
Correlation (↑)	0.466	0.449	1
Recall (↑)	0.568	0.378	-
Precision (↑)	0.072	0.108	-

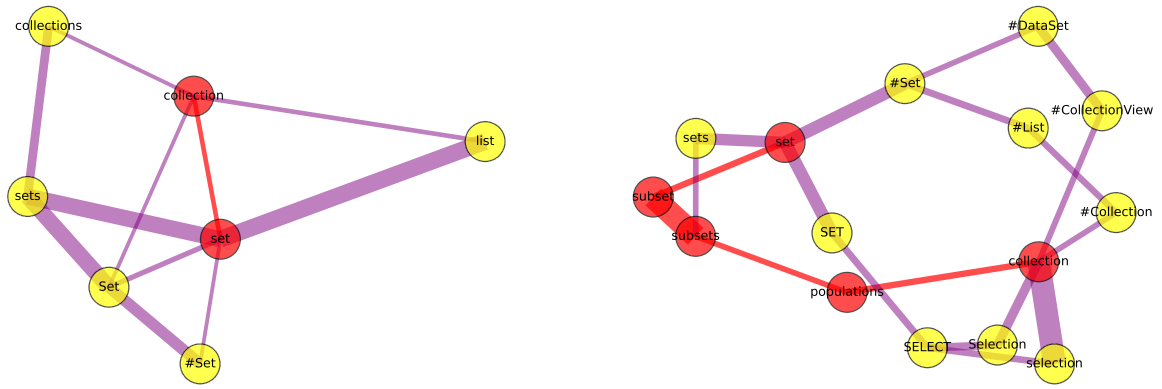
Table 3: Statistics for Llama2-7B, Llama2-70B, and GT across four dimensions. "#Component" represents the number of connected components, and "#Single" indicates the number of nodes with zero degree. In the bottom section, we report degree correlation, recall, and precision.

Overall, the model-constructed networks contain more edges and fewer connected components and isolated nodes than those built by human ex-



(a) Llama2-7B (b) Llama2-70B

Figure 6: Shortest paths between “bar” and “library” in Llama2-7B and Llama2-70B. Edge width indicates weight; red highlights mark the shortest path. The 70B model shows a more complex structure involving similar forms, multilingual links, and disambiguation clusters.



(a) Llama2-7B (b) Llama2-70B

Figure 7: Shortest paths between “collection” and “set” in Llama2 models. Red highlights indicate the shortest path. The 70B model shows a more logical semantic transition from singular to plural concepts.

perts. Expert-constructed graphs are based on corpus co-occurrence across at least three languages, a process especially difficult for low-resource languages. In contrast, model embeddings produce much denser graphs. Notably, Llama2-70B yields a sparser network than Llama2-7B. Compared to the ground truth (GT), the automatic networks show moderate correlation and coverage (recall), indicating that the lexico-semantic network partially captures universal cross-linguistic patterns of semantic relatedness in polysemy. However, precision is lower due to the larger number of edges, suggesting embeddings can serve as a preliminary space for linguists to further refine.

6 Conclusion

This paper investigates the construction of lexico-semantic networks using input embeddings from large language models (LLMs). We analyze and compare the network properties of LLMs with different scales across three scenarios. Our findings show that the lexico-semantic network can be effectively constructed from embeddings, which exhibit a small-world clustering effect. Additionally, models with more parameters tend to explore longer and more complex paths between concepts, partially supporting the “scaling law” (Kaplan et al., 2020). This study also provides an efficient approach to constructing conceptual spaces, potentially benefiting fields such as language typology and cognitive science.

7 Limitations

We acknowledge several limitations in our work. First, we mainly concentrate on the monosemous words. However, words can be ambiguous, particularly for homonyms that encompass multiple unrelated meanings. Second, our evaluation is limited to two models, Llama2-7B and Llama2-70B. Results may differ with models of different architectures or parameter scales. Additionally, we focus only on input embeddings and do not explore the properties of output embeddings, which may also capture individual word representations. Finally, the length of the shortest path in the lexico-semantic network is not a definitive metric for embedding quality, and we plan to explore more sophisticated metrics to better reflect the characteristics of these spaces.

8 Ethics Statement

We do not foresee immediate ethical concerns arising from our research. However, there may be unintended biases in the connections between concepts, such as those involving gender and job ranks. These biases may stem from biased embeddings (Bordes et al., 2013; Bolukbasi et al., 2016).

References

Réka Albert, Hawoong Jeong, and Albert-László Barabási. 1999. Internet: Diameter of the world-wide web. *Nature*, 401:130–131.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Gemma Boleda. 2020. Distributional semantics and linguistic theory. *Annual Review of Linguistics*, 6(1):213–234.

Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2787–2795.

Leyla Roksan Caglar. 2021. *Conceptual Spaces in the Brain: An Exploration of Structure, Shape, and Organization*. Ph.D. thesis, Rutgers The State University of New Jersey, Graduate School-Newark.

William Croft. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press, Oxford.

Johannes Dellert. 2024. Causal inference of diachronic semantic maps from cross-linguistic synchronic polysemy data. *Frontiers in Communication*, 8:1288196.

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.

John Firth. 1957. A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*, pages 10–32.

Matthew Freestone and Shubhra Kanti Karmaker Santu. 2024. Word embeddings revisited: Do llms offer something new. *arXiv preprint arXiv:2402.11094*.

Rui Guo. 2012. The typology of adjectives and the grammatical status of chinese adjectives. *Chinese Language Learning*, (5).

P. Gärdenfors. 2000. *Conceptual Spaces: The Geometry of Thought*. MIT Press.

P. Gärdenfors. 2014. *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. MIT Press.

Chi Han, Jialiang Xu, Manling Li, Yi Fung, Chenkai Sun, Nan Jiang, Tarek Abdelzaher, and Heng Ji. 2024. Word embeddings are steers for language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16410–16430, Bangkok, Thailand. Association for Computational Linguistics.

Zellig S Harris. 1954. Distributional structure. *Word*, 10(2-3):146–162.

Martin Haspelmath. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello, editor, *The new psychology of language*, volume 2, pages 211–243. Lawrence Erlbaum, Mahwah, NJ.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Zhu Liu, Cunliang Kong, Ying Liu, and Maosong Sun. 2024. Fantastic semantics and where to find them: Investigating which layers of generative LLMs reflect lexical semantics. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 14551–14558, Bangkok, Thailand. Association for Computational Linguistics.

Zhu Liu, Cunliang Kong, Ying Liu, and Maosong Sun. 2025. A top-down graph-based tool for modeling classical semantic maps: A case study of supplementary adverbs. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 4567–4576, Albuquerque, New Mexico. Association for Computational Linguistics.

Andrej Malchukov, Martin Haspelmath, and Bernard Comrie. 2007. Ditransitive constructions: A typological overview. In *Paper for the Conference on Ditransitive Constructions*, Max Planck Institute for Evolutionary Anthropology, Leipzig.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 3111–3119.

Stanley Milgram. 1967. The small world problem. *Psychology Today*, 2:60–67.

George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

M. Moullec and I. Douven. 2025. Cheaper spaces. *Minds & Machines*, 35(6).

R. M. Nosofsky. 1986. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115:39–57.

Loïc-Michel Perrin. 2010. Polysemous qualities and universal networks, invariance and diversity. *Linguistic Discovery*, 8:1–22.

Erika Petersen and Christopher Potts. 2023. Lexical semantics with large language models: A case study of English "break". In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 490–511, Dubrovnik, Croatia. Association for Computational Linguistics.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.

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*.

Peter D. Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37:141–188.

Ivan Vulić, Edoardo Maria Ponti, Robert Litschko, Goran Glavaš, and Anna Korhonen. 2020. Probing pretrained language models for lexical semantics. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7222–7240, Online. Association for Computational Linguistics.

Duncan J. Watts and Steven H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442.

Ying Zhang. 2017. Semantic map approach to universals of conceptual correlations: a study on multifunctional repetitive grams. *Lingua Sinica*, 3(1):7.

Joost Zwarts. 2010. Semantic map geometry: Two approaches. *Linguistic Discovery*, 8(1):377–395.

A Appendix

A.1 Common Concepts in Scenario 1

Table 4 lists specific concepts from different semantic groups, with each group containing ten common concepts.

A.2 Conceptual Spaces in Scenario 3

Scenario 3 presents a human-annotated conceptual space for qualitative words, as shown in Figure 8. Each concept is represented by an English word. Nodes are connected if a pair of concepts co-occur as a polysemous word in at least three languages. Nodes marked in red represent federative words, indicating a shared concept with a higher degree.

A.3 Degree Distribution

We show the distribution of node degrees for spaces generated by two models, Llama2-7B and Llama2-70B. Figure 9(a) displays the unweighted degree distribution, while (b) shows the weighted distribution. The results indicate that the 70B model has a less pronounced long-tail distribution, with more nodes having relatively larger degrees.

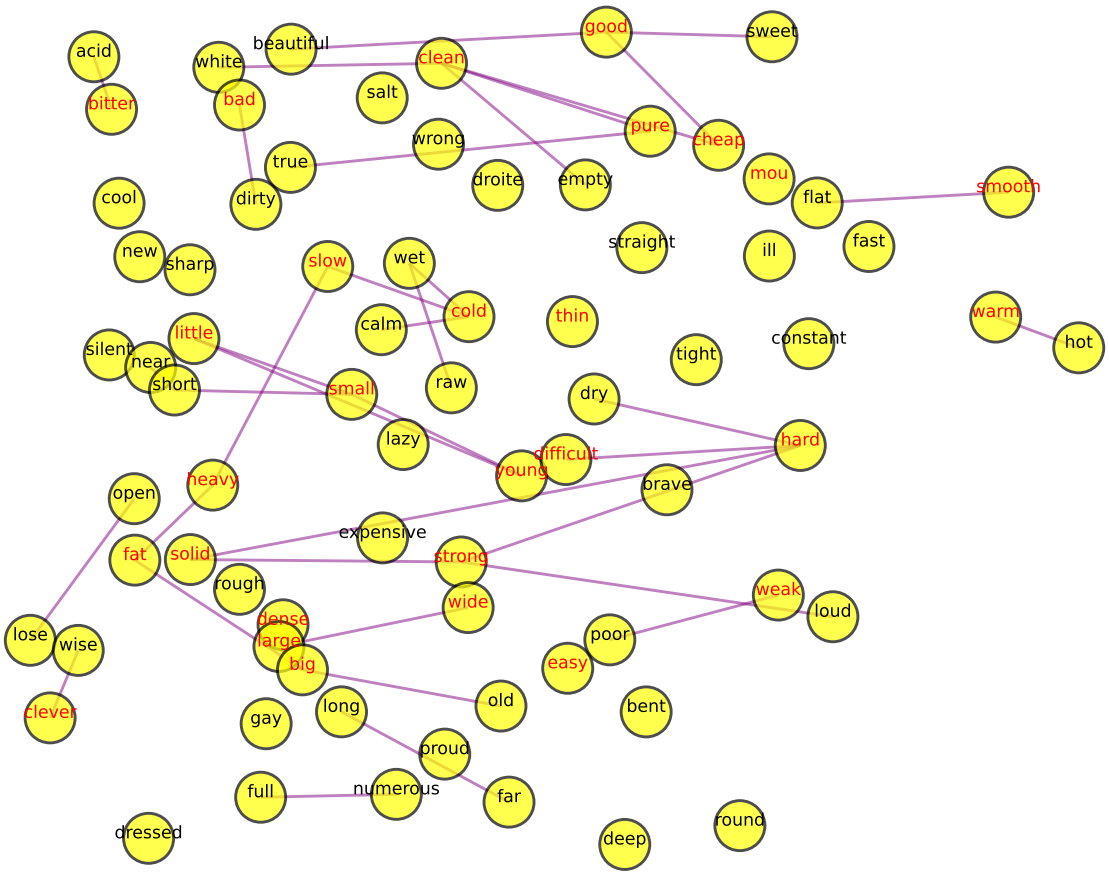


Figure 8: A semantic map for the domain of qualitative words, with federative notions which have a higher degree highlighted in red.

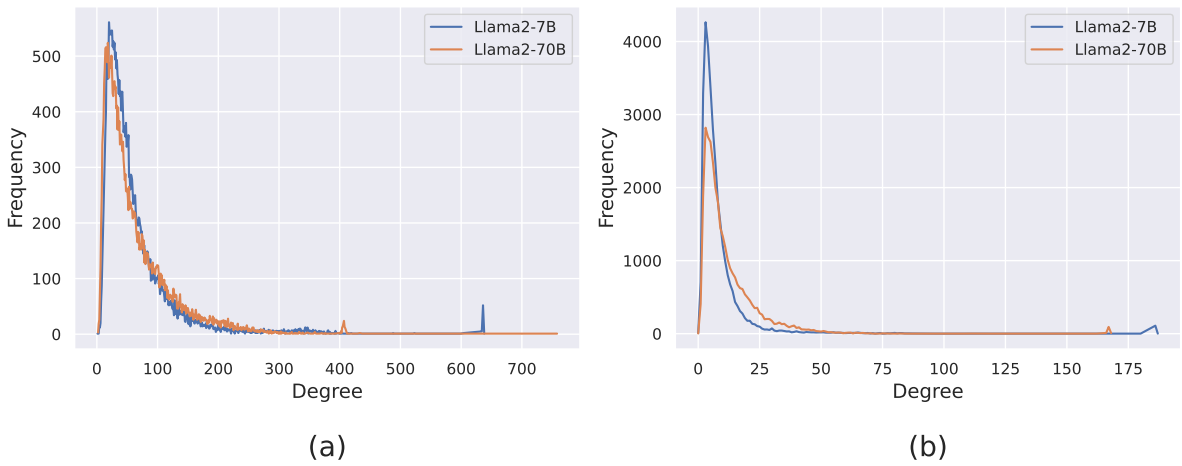


Figure 9: Comparison of the degree distribution for both models: (a) unweighted degree and (b) weighted degree.

Semantic Group	Words
NUMBER	one, two, three, four, five, six, seven, eight, nine, ten
NAME	Alice, Bob, Carol, Dave, Francis, Grace, Hans, Ivan, Zach, Mike
MONTH	January, February, March, April, May, June, July, August, September, October
COLOR	red, orange, yellow, green, blue, brown, black, white, grey, gray
CITY	Taiwan, York, Cambridge, Oxford, Berlin, Paris, Washington, Rome, Tokyo, Toronto
NATION	China, America, England, UK, Germany, France, USA, Italy, Japan, Spain
PLACE	factory, concert, museum, library, bar, zoo, park, theater, hospital, church
HUMAN	female, male, man, woman, human, boy, girl, elder, gentleman, guys
FURNITURE	chair, desk, table, bed, cabinet, computer, lamp, mirror, house, room
RANDOM	conscious, distance, measure, almost, paste, sun, friend, other, waste, tongue

Table 4: Specific words from different semantic groups.