A simple clustering approach to map the human brain's cortical semantic network organization during task

Yunhao Zhang , Shaonan Wang , Nan Lin , Lingzhong Fan , Chengqing Zong

PII: S1053-8119(25)00098-9
DOI: https://doi.org/10.1016/j.neuroimage.2025.121096
Reference: YNIMG 121096

To appear in: *NeuroImage*

Received date: 21 August 2024
Revised date: 5 February 2025
Accepted date: 18 February 2025

Please cite this article as: Yunhao Zhang , Shaonan Wang , Nan Lin , Lingzhong Fan , Chengqing Zong , A simple clustering approach to map the human brain's cortical semantic network organization during task, *NeuroImage* (2025), doi: https://doi.org/10.1016/j.neuroimage.2025.121096

Highlights

- We propose a novel clustering method for partitioning large-scale brain networks based on specific cognitive functions, focusing on semantic representation as the target function.
- Our method reveals distinct brain network organizations during cognitive tasks, identifying seven unique semantic networks, highlighting differences from resting-state networks.
- A strong correlation is observed between the stability of the identified semantic networks and their semantic representation capabilities, providing new insights into the functional organization of the brain.
- Our findings reveal that certain semantic networks integrate information from both linguistic and visual sources, while others predominantly rely on visual inputs.

A simple clustering approach to map the human brain's cortical semantic network organization during task

Yunhao Zhang[1,2], Shaonan Wang[1,2*], Nan Lin[3,4*], Lingzhong Fan[2,5], Chengqing Zong[1,2]

[1] State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, CAS, Beijing, China

[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China

[3] CAS Key Laboratory of Behavioural Sciences, Institute of Psychology, Beijing, China

[4] Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

[5] Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China

Correspondence should be addressed to Shaonan Wang (shaonan.wang@nlpr.ia.ac.cn) and Nan Lin (linn@psych.ac.cn)

*Corresponding author

Abstract

Constructing task-state large-scale brain networks can enhance our understanding of the organization of brain functions during cognitive tasks. The primary goal of brain network partitioning is to cluster functionally homogeneous brain regions. However, a brain region often serves multiple cognitive functions, complicating the partitioning process. This study proposes a novel clustering method for partitioning large-scale brain networks based on specific cognitive functions, selecting semantic representation as the target cognitive function to evaluate the validity of the proposed method. Specifically, we analyzed functional magnetic resonance imaging (fMRI) data from 11 subjects, each exposed to 672 concepts, and correlated this with semantic rating data related to these concepts. We identified distinct semantic networks based on the concept comprehension task and validated the robustness of our network partitioning through multiple methods. We found that the semantic networks derived from multidimensional semantic activation clustering exhibit high reliability and cross-semantic model consistency (semantic ratings and word embeddings extracted from GPT-2), particularly in networks associated with high semantic functions. Moreover, these semantic networks exhibits significant differences from the resting-state and task-based brain networks obtained using traditional methods. Further analysis revealed functional differences between semantic networks, including disparities in their multidimensional semantic representation capabilities, differences in the information modalities they rely on to acquire semantic information, and varying associations with general cognitive domains. This study introduces a novel approach for analyzing brain networks tailored to specific cognitive functions, establishing a standard semantic parcellation with seven networks for future research, potentially enriching our understanding of complex cognitive processes and their neural bases.

Keywords: Task-state large-scale brain networks, Semantic rating, GPT, Concept, fMRI

1. Introduction

In recent years, one of the most significant advancements in neuroscience has been the discovery of large-scale brain networks, where distant brain regions exhibit synchronized neural activities, collectively forming functionally homogeneous networks (Eickhoff et al., 2018; Petersen and Sporns, 2015; Power et al., 2010). This discovery has substantially enhanced our understanding of the brain's functional organization (Bressler and Menon, 2010; Pessoa, 2014). The current delineation of large-scale brain networks is primarily based on findings from resting-state fMRI studies, which identify multiple networks by measuring the correlation of neural activities in distant brain regions using resting-state functional connectivity (RSFC) patterns (Ji et al., 2019; Power et al., 2011; Thomas Yeo et al., 2011). The prominence of these resting-state networks is such that many studies investigating task-state brain representation patterns continue to rely on networks defined by resting-state fMRI data (Lin et al., 2018a, 2024; Sun et al., 2024; Zhang et al., 2024). In addition to resting-state networks, another popular method for brain network parcellation is based on structural information, such as white matter connectivity (Fan et al., 2016) and cortical morphology (Destrieux et al., 2010). These structurally defined networks are also widely employed in task-state fMRI research (Bruurmijn et al., 2017; Fischer et al., 2021; Li et al., 2019; Schneider et al., 2022; Y. Wang et al., 2024). However, in many cognitive neuroscience studies, researchers often focus on a specific cognitive function. The question they aim to address through brain network analysis is the brain-network organization of a specific cognitive function in the brain. Network partitions derived from resting-state functional connectivity (RSFC) or structural information may not be sufficient for such requires. This is because stronger RSFC or structural connectivity between regions does not necessarily indicate greater functional homogeneity for the target cognitive function, nor does weaker connectivity always correlate with increased functional heterogeneity.

One potential solution to the aforementioned issue is to partition brain networks based on task fMRI data. Experimental tasks are typically designed to examine the influence of specific cognitive functions on brain activation, making this approach suitable for investigating the brain-network organization of cognitive functions in relation to particular tasks. Studies have found that, under task conditions, brain networks undergo reorganization compared to the resting state. For example, Rolinski et al. (2020) compared the brain network distribution under resting state and a language task and found a moderate overlap between the networks in these two states, with a notable shift from left to bilateral dominance at rest, suggesting a more distributed organization in resting networks. Similarly, Doucet et al. (2017) observed reduced lateralization in language

networks at rest. Jackson et al. (2016) reported enhanced connectivity from the anterior temporal lobe to occipital and frontal cortex regions during semantic tasks, highlighting task-specific network expansions. Cole et al. (2021) demonstrated that task-state functional connectivity (TSFC) predicts activations across various tasks and cortical areas more accurately than RSFC. Moreover, some studies have attempted to predict task-state fMRI activations from resting-state fMRI data at the individual level using general linear models and neural network methods (Cohen et al., 2020; Jones et al., 2017; Ngo et al., 2022a; Niu et al., 2021; Tavor et al., 2016). These studies found that although neural network methods outperform general linear models in prediction accuracy, the overall prediction accuracy remains relatively low (Cohen et al., 2020; Deco et al., 2015; Jones et al., 2017). To capture the functional reorganization and dynamic changes of brain networks during cognitive tasks, researchers have proposed various methods for analyzing the organization of brain networks based on task-related BOLD signals, including: clustering algorithms (Salehi et al., 2020), machine-learning algorithms (Glasser et al., 2016), and meta-analysis (Dockès et al., 2020; Laird et al., 2005; Ngo et al., 2022b; Yarkoni et al., 2011). In contrast to employing resting-state or structural brain network partition, these methodologies offer more optimized and adaptable choices for network analyses in task-based fMRI research (Ngo et al., 2019; Yeo et al., 2015).

However, while task-based brain network analysis methods are more task-specific, they are still unable to precisely capture the brain-network organization associated with target specific cognitive functions. The execution of experimental tasks often necessitates the coordination of the target cognitive function with various non-target cognitive functions, and these functions often jointly influence the neural activity of a brain region (Hein and Knight, 2008; Mattheiss et al., 2018). Conventional task-based parcellation methods often fail to distinguish between neural activities related to target and non-target cognitive components. Therefore, these methods are often influenced by both target and non-target cognitive functions (as well as potential physiological attributes), which prevents them from maximizing the functional homogeneity of the target cognitive function within the identified networks (Kuhnke et al., 2023; Lin et al., 2020; G. Zhang et al., 2023). For instance, consider a task where the neural activities of brain regions X, Y, and Z are influenced by three cognitive functions: A, B, and C, each contributing equally to the neural signals. For function A, X and Y are functionally homogeneous but different from Z. For functions B and C, X and Y are functionally opposite but similar to Z. If we consider the overall activity of these regions during the task, X and Z would be grouped into the same network due to their higher overall homogeneity. However, if the researcher is only interested in function A, then

X should be grouped with Y rather than Z, because X and Y are functionally homogeneous regarding function A.

To meet the aforementioned requirements, we propose a method for partitioning large-scale brain networks based on specific cognitive functions. The core concept involves isolating information related to the target cognitive function from neural signals and deconstructing it into multiple components or dimensions. Then brain parcels are clustered based on their functional attributes across these components or dimensions to form large-scale brain networks.

In this study, we chose semantic representation function as the target cognitive function to evaluate the validity of the proposed method. There are three reasons for this choice. First, semantic representation is a fundamental and crucial cognitive function that underpins various important abilities, including language comprehension and production, object recognition and classification, and understanding everyday events (Frisby et al., 2023; Kumar, 2021; S. Wang et al., 2024). Second, extensive research has identified effective multi-dimensions semantic models of semantic representation, such as interpretable semantic dimension ratings (Anderson et al., 2017; Fernandino et al., 2015, 2022; Tong et al., 2022; Y. Zhang et al., 2023b) and word embeddings from computational language models (Caucheteux et al., 2023; Schrimpf et al., 2021; A. Y. Wang et al., 2023). Third, studies have found that semantic representation is distributed across a wide range of brain regions, with functional heterogeneity among these regions (Fernandino et al., 2016; Huth et al., 2016, 2012), suggesting the likely presence of multiple subnetworks in the brain supporting semantic representation.

Specifically, we analyzed fMRI data from a concept comprehension task, in which 11 participants thought about 672 individual concepts. We used interpretable semantic dimension ratings to disentangle multidimensional semantic-related information from the fMRI data and then partitioned the cortical network based on the semantic representation functions of brain regions. We then assessed the reliability of our method, and examined the validity of our method by analyzing the similarities between semantic-network partitions obtained using different semantic models (i.e., interpretable semantic dimension ratings and GPT-2) and between semantic-network partitions and brain networks obtained using traditional approaches (i.e., resting-state RSFC and modularity analysis). Finally, we examined the functional differences between semantic networks in their multidimensional semantic representation capabilities, semantic acquisition sources, and associations with general cognitive domains.

2. Methods

Figure 1 provides an overview of the study design, which encompasses four key steps: data processing, semantic network partition, assessing quality of semantic network partition, and examination of the cognitive functions of the partitioned semantic networks. Each of these components is discussed in detail below.

2.1 Data processing

To construct the semantic networks, we utilized two recently published datasets: the fMRI dataset for concept representation with semantic feature annotations (CRSF) (S. Wang et al., 2022b) and the six semantic dimension database (SSDD) (S. Wang et al., 2023). In this subsection, we detail the fundamental characteristics of the two datasets, outline our processing procedures for each (see Figure 1a).

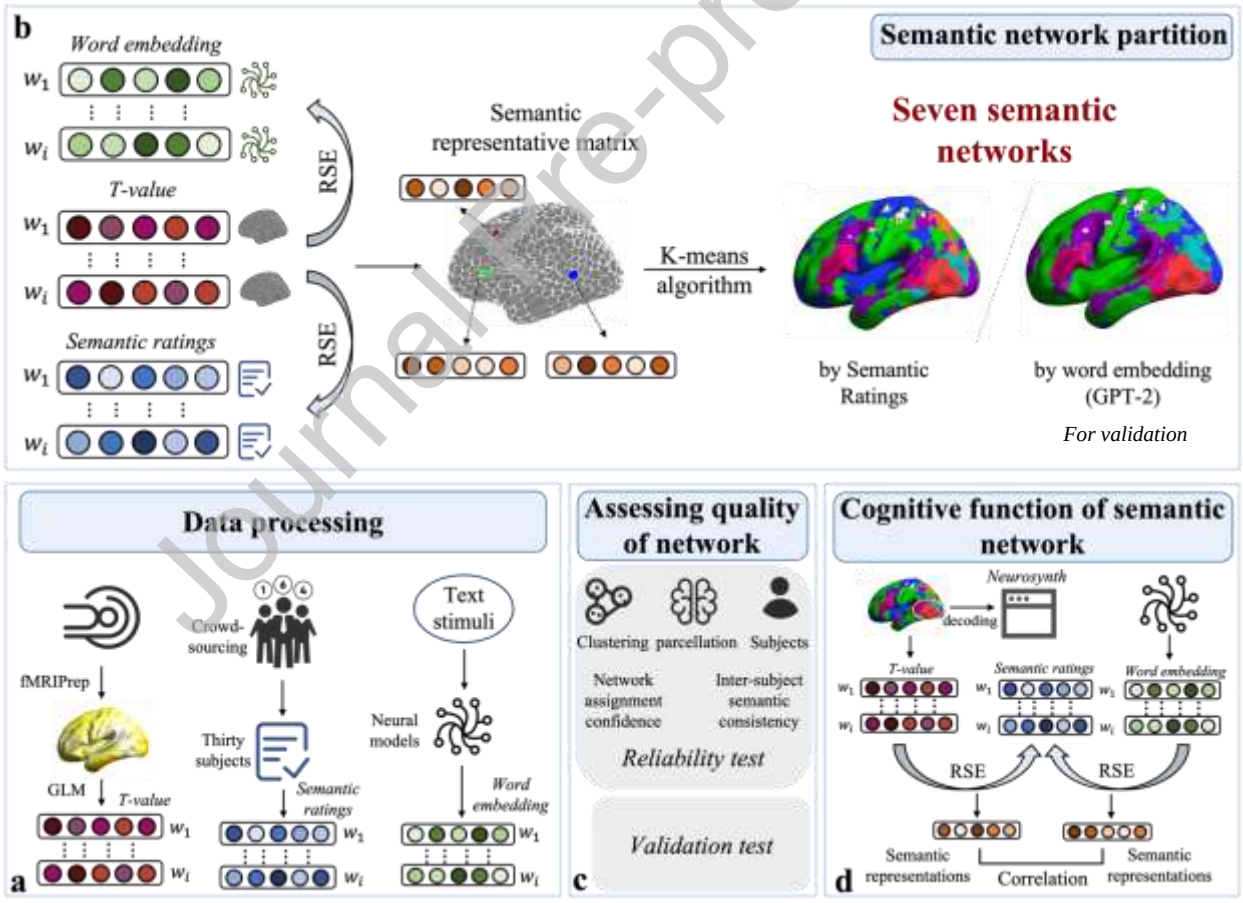

Fig 1. Overview of the study design. The procedure includes four parts: A) Data processing: initial fMRI data was processed using fMRIPrep, and t-value images for each concept were obtained using the general linear model. A 59-dimensional semantic ratings for each concept was

constructed by incorporating CRSF and SSDD datasets. B) Semantic network partition: The t-value image was divided regionally according to 1,000 independently identified parcels for each subject and each concept. Then, t-value representations of each parcel were mapped to semantic ratings or word embeddings using representational similarity encoding, resulting in a 1,000 × 59 semantic representation matrix for each subject. After averaging across subjects, the k-means algorithm was applied to create a semantic cortical partition. C) Assessing quality of semantic network partition was performed from two perspectives: reliability test and validation test. D) Examination of the cognitive functions of the partitioned semantic networks from three perspectives: their multidimensional semantic-representation capabilities, sources of semantic information acquisition and involvement in general cognitive domains.

2.1.1 The fMRI data from the CRSF

For detailed information on the fMRI portion of the CRSF dataset, please refer to S. Wang et al. (2022b) This dataset initially included 18 participants (8 females, mean age 23.83 ± 2.4 SD), though data from 7 participants were excluded due to incomplete participation (average 1.43 visits ± 0.73 SD). Our analysis focuses on the 11 right-handed individuals who considered 672 unique concepts from both concrete and abstract categories, sourced from the Synonymy Thesaurus published by Harbin Institute of Technology (HITST).

During fMRI scanning, participants were instructed to attentively read the presented words and consider their related concepts in conjunction with the accompanying images. Specifically, at the beginning of each run, the instruction "The experiment is about to start; please pay attention" was displayed on the screen, followed by a 2-second fixation period. Each stimulus was then presented for 3 seconds, followed by a 2-second fixation period. The fMRI sessions were split into four visits for participants sub01–sub05 and six visits for participants sub06–sub11. Within each scanning session, the 672 words were divided into either four sets of 168 words (for sub01–sub05) or six sets of 112 words (for sub06–sub11), and distributed across 12 runs. Each participant viewed each word six times, each with a different picture. It took two runs to complete a single repetition of all 168 or 112 words (i.e., 84 or 56 words per run).

Data collection employed a 3T GE Discovery MR750 scanner with a 32-channel phased-array head coil at the Magnetic Resonance Imaging Research Center of the Institute of Psychology of the Chinese Academy of Sciences (IPCAS). We acquired T1-weighted structural images in 176

sagittal slices (1.0 mm isotropic voxels) and functional BOLD signals using gradient-echo EPI in 42 near-axial slices (3.0 mm isotropic voxels, TR 2000 ms, TE 30 ms, flip angle 70 degrees).

Automated pre-processing of the images was performed using fMRIPrep (Esteban et al., 2019), which included initial discarding of the first 5 volumes of each functional run, slice timing correction, 3-D motion correction, and standard space resampling. T1-weighted images underwent defacing, manual inspection of cortical segmentations, and normalization. Specifically, T1-weighted images were segmented into different tissue types; the resulting gray matter probabilistic images were coregistered to the mean functional image in the native space, resliced to the spatial resolution of functional images, and obtain the gray mask of each subject. The forward and inverse deformation fields of each subject's native space to the Montreal Neurological Institute (MNI) space were also obtained at this step.

After preprocessing, we estimated the fMRI single-word responses. For each participant, we applied a general linear model (GLM) to the fMRI data in standard space to obtain word-level neural activation patterns. The GLM included onset regressors for each of the 672 words, six motion parameters, and a constant regressor for each run, with convolution through the canonical hemodynamic response function (HRF) and a high-pass filter set at 128 s. For each word, this process generated six beta-value images reflecting neural activation patterns. Next, for each word, we performed a one-sample t-test on the six beta-value images (with a comparison to zero) to obtain the corresponding t-value image, which reflects stable word-level neural activation pattern. The use of the t-map in this analysis is motivated by the fact that, in activation pattern analysis, it is crucial to standardize the activation data for each voxel. A one-sample t-test achieves this standardization by comparing the activation values at each voxel against the null hypothesis (zero). Finally, following previous studies (Anderson et al., 2016; Fu et al., 2023), we only consider voxels in the gray matter mask for subsequent analysis.

2.1.2 The semantic ratings from the CRSF and the SSDD

The semantic rating-based method is primarily grounded in the multidimensional semantic model developed by Binder et al. (2016). Drawing on neuroscience research related to semantics, Binder et al. (2016) proposed a semantic model that encompasses 65 semantic dimensions across 14 domains, along with guidelines for their assessment. Subsequent research has validated the semantic rating-based model's effectiveness in explaining semantic-related behaviors and brain activation patterns (Anderson et al., 2017; Fernandino et al., 2022, 2015; Tong et al., 2022; Y.

Zhang et al., 2023b). Moreover, the rating-based model is known for its high interpretability, which makes it an ideal choice for the current study. In this research, we utilized a 59-dimensional Chinese semantic ratings derived from two datasets: the CRSF dataset and the SSDD dataset.

We direct the reader to S. Wang et al. (2022b) for details on the semantic ratings of the CRSF dataset. The dataset comprised 54 semantic features for 672 concepts across 14 domains (vision, somatic, audition, gustation, olfaction, motor, spatial, temporal, causal, social, cognition, emotion, drive, attention). Each concept was evaluated on a 1-7 scale by crowd-sourced experiments involving 30 unique raters per experiment. A total of 126 participants (72 females, aged 20-25) contributed, completing tasks based on their ability to pass quality assessments. These 54 dimensions are derived from the multidimensional semantic model proposed by Binder et al. (2016). Eleven of the original 65 features in Binder's model were excluded due to high correlations (Pearson correlation > 0.8) with other features.

Additionally, our study utilized the subjective rating dataset from the SSDD, which includes subjective ratings for 17,940 Chinese words, over the computational extension dataset, which contains ratings for over 1.4 million Chinese and 1.5 million English words. The current study used the subjective rating dataset because it showed higher validity than the computational extension dataset and includes the semantic ratings of 672 words used in the CRSF. It focuses on 6 semantic dimensions: vision, motor, socialness, emotion, space, and time. Except for the vision dimension, the other five dimensions are not rated in the CRSF. Notably, emotion ratings range from -3 to 3 (often termed valance ratings), whereas the other dimensions use a 1-7 scale. Following conventions in neuroimaging research on emotional semantics (see Arioli et al. (2021)), we used absolute values of the emotional ratings to differentiate between valenced and neutral words in neural representations.

Notably, the "Vision" feature was rated in both datasets, showing significant overlap and a high correlation between them. Consequently, we retained the "Vision" feature from the SSDD dataset and merged the semantic ratings from both sources, creating a composite set of 59-dimensional semantic ratings for each concept. Each of the 59 dimensions in this composite rating set has clear and well-defined semantic connotations, as detailed in Supplementary Table S1. Supplementary Figure S1 illustrates three examples of the semantic ratings. As expected, more concrete concepts,

such as "Airplane" and "Sea," received higher ratings on sensory and motor domains, whereas abstract concepts like "Love" were rated higher in abstract domains.

2.2 Semantic network partition

As shown in Figure 1b, to achieve a large-scale semantic network organization, we utilized a cortical parcellation to sample fMRI data at the regional level (Ji et al., 2019). The recently-developed cortical parcellation by Schaefer et al. (2018) includes 500 symmetric cortical parcels per hemisphere. This parcellation, defined by surface vertices, is considered more accurate than previous versions due to the use of a Gaussian Markov Random Field (gwMRF) model, which integrates local gradient and global similarity approaches. Each parcel varies in size and shape, aligning functional and anatomical borders across multiple imaging modalities. For each subject and each concept, we divided the t-value image (detailed in Section 2.1.1) regionally according to 1,000 independently identified parcels. This process resulted in v -dimensional t-value representations (where v equals the number of voxels in the parcel) for each parcel. We selected a higher-resolution Schaefer atlas rather than a lower-resolution version (e.g., 100 or 200 symmetric cortical parcels per hemisphere) to more effectively partition voxels with distinct semantic information distributions into separate parcels. Specifically, voxels that are spatially proximal tend to exhibit functional homogeneity; consequently, smaller parcels are likely to demonstrate greater internal homogeneity in terms of semantic representation functions. In contrast, using a lower-resolution atlas increases the risk that voxels with highly heterogeneous semantic representation functions are grouped within the same parcel. Therefore, employing a higher-resolution Schaefer atlas maximizes the homogeneity of semantic representation functions within brain networks.

The v-dimensional t-value representations of each parcel (detailed in Section 2.1.2) contains activations triggered by both semantic and non-semantic components (such as attention and memory retrieval). To eliminate noise and semantically irrelevant information in brain data, we mapped v-dimensional t-value representations of each parcel to 59-dimensional semantic ratings. This multi-dimensional decoding method is commonly used both to remove uninterested information from computational language model representations in natural language processing (Oota et al., 2024; Toneva et al., 2022) and to isolate brain activity related to the specific NLP task from fMRI cognitive signals (Luo et al., 2022; Y. Zhang et al., 2023a). For each parcel, we mapped the t-value representations to each dimension of the semantic ratings to generate predicted semantic rating vectors using representational similarity encoding (RSE), as detailed in

Section 2.5. Subsequently, for each parcel, we obtained a 59-dimensional semantic representation, consisting of the Pearson correlation coefficients between the actual and predicted semantic rating vectors. This procedure was repeated for 1,000 parcels, resulting in a 1,000 × 59 semantic representation matrix for each subject. A group semantic representation matrix was subsequently created by averaging these matrices across all subjects in the cohort.

We then applied the k-means clustering algorithm to segment 1,000 brain parcels into major cortical semantic networks based on the group-level semantic representative matrix. The optimal number of clusters, k, was determined using multiple methods, consisting of the elbow method, silhouette coefficient and cross-subject stability. The elbow method assesses the intra-cluster sum of squares (inertia) across different k values to identify significant changes in the rate of inertia reduction:

$$\text{Inertia}(k) = log(\sum_{i=1}^{n} \min_{\mu_j \in S_k} (||x_i - \mu_j||^2)) \quad (1)$$

Here, n represents the number of parcels, x_i is the i-th parcel, μ_j is the cluster center within the set S_k.

The silhouette coefficient (SC) measures each data point's similarity to others within the same cluster (cohesion) and its dissimilarity to data points in the nearest different clusters (separation):

$$\text{SC} = -log(\frac{1}{n}\sum_{i=1}^{n} \frac{v_i - w_i}{max(w_i, v_i)}) \quad (2)$$

where w_i denotes the average intra-cluster distance for data point i, reflecting cluster cohesion, and is computed as:

$$w_i = \frac{1}{|S|-1} \sum_{j \in S, j \neq i} \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2} \quad (3)$$

Here, S represents the cluster containing point i, and the Euclidean distance serves as the distance metric. v_i represents the average distance from data point i to the nearest point in a different cluster S', capturing cluster separation, and is calculated as:

$$v_i = \min_{S' \neq S} \frac{1}{|S'|} \sum_{j \in S'} \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2} \quad (4)$$

Here, S' denotes any cluster other than S. In the formula 2, a lower silhouette coefficient indicates more effective clustering, reflecting a clearer distinction between clusters.

Cross-subject stability was calculated by averaging the pairwise Euclidean distances between the clustering centers derived from each subject's semantic representative matrix for various cluster numbers k. Specifically, the mean distance $D^{(k)}$ was determined by aggregating the distances between corresponding cluster centers across all pairs of subjects, for each k. This aggregation was mathematically represented as:

$$D^{(k)} = \frac{1}{\binom{N}{2}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{1}{k} \sum_{m=1}^{k} \| \boldsymbol{S}_{i,k}^{(m)} - \boldsymbol{S}_{j,k}^{(m)} \|_2 \quad (5)$$

where $\boldsymbol{S}_{i,k}^{(m)}$ and $\boldsymbol{S}_{j,k}^{(m)}$ denote the m-th cluster center of subjects i and j respectively, and N is the total number of subjects. This methodological approach facilitated the identification of the most stable clustering number k, which was characterized by the smallest mean distance across subjects, thus reflecting the highest consistency in clustering patterns among the subjects.

2.3 Assessing quality of semantic network partition

In the previous section, we constructed semantic network partitions using both semantic ratings and word embeddings. In this section, we conducted reliability and validation tests on the network partition method to assess its stability and scalability (See Figure 1c).

2.3.1 Reliability test

As shown in Figure 1c, we assessed the reliability of our network partition at both the global and local levels. At the global level, we conducted validation analyses employing different clustering algorithms, various cortical parcellation methods, and different subject sampling subsets. These factors are crucial in determining the final semantic network partition. Initially, we explored the effect of various clustering methods on network partitioning by re-clustering the semantic representation matrix using multiple algorithms, including k-means clustering, spectral clustering, hierarchical clustering, and Gaussian mixture models. Subsequently, to evaluate the influence of cortical parcellations, we utilized the human brainnetome atlas (BN atlas) (Fan et al., 2016) and Schaefer's parcellations (Schaefer et al., 2018), as outlined in Section 2.2. The BN atlas,

comprising 210 cortical parcels, offers a detailed, cross-validated network map with anatomical and functional connectivity data. Schaefer's parcellations generated parcels ranging from 400 to 1,000 to accommodate various applications, all of which were included in our analysis. To determine the impact of subject variability on network partitioning, we repeatedly selected half of the subjects at random to replicate the analysis process. Finally, we calculated the Pearson correlation for semantic networks generated by different methods to quantify consistency.

At the local level, we implemented several quantitative measures to rigorously evaluate the final semantic network partition and validate parcel assignments. First, we calculated a network assignment confidence score for each parcel to express the certainty of its assignment to a particular network (Wang et al., 2015). This confidence score was determined by the difference between the assigned network's correlation value and the out-of-network correlation values for a parcel i:

$$C_i = \frac{\sum r_{p\,i,j}}{m_j} - \frac{\sum r_{p\,i,h}}{m_h} \quad (6)$$

where C_i is the network assignment confidence score for parcel i (one of 1,000 brain parcels), $r_{p\,i,j}$ represents the Pearson correlation coefficient between the semantic representations of parcel i and another parcel j within the same network, m_j is the total number of parcels within parcel i's network. $r_{p\,i,h}$ is the Pearson correlation coefficient between the semantic representations of parcel i and parcel h outside parcel i's network, and m_h denotes the count of parcels outside parcel i's network. If a parcel's semantic representation is very similar to that of the other parcels in its assigned network, the confidence score will be high, but if it is also similar to other networks, the confidence score will be low.

Then, we evaluated inter-subject semantic consistency to gauge the consistency of each parcel's semantic representation across subjects (Ji et al., 2019). For each parcel, we calculated the Pearson correlation between the semantic representations of one subject and all others, resulting in an 11 × 11 matrix (the number of subjects × the number of subjects). The mean pairwise similarity score S was obtained by averaging the matrix values and used as the inter-subject semantic consistency score. This score quantifies the degree of semantic consistency among subjects for each parcel. A higher score indicates higher inter-subject semantic consistency, whereas a lower score reflects lower consistency across subjects.

Once these quality metrics were calculated, each parcel was evaluated for reassignment. Specifically, parcels with a confidence score below 0 were considered for reassignment. For these parcels, we calculated the confidence scores for assignment to each available semantic network and reassigned the parcel to the network with the highest confidence score. This reassignment process ensured that each parcel was assigned to the most appropriate network based on quantitative assessments. Except for parcels in network 3, less than 3% of cortical patches in other networks were reallocated.

In Network 3, parcels were not subjected to reassignment. Notably, over 70% of the parcels within this network exhibit confidence scores below zero, resulting in an overall negative confidence score for Network 3. Two primary factors may account for this negative confidence score. First, the homogeneity of parcels within Network 3 is low; the application of a hard clustering analysis method (i.e., K-means) likely aggregated parcels that were difficult to assign to other categories into a single network. Second, regarding semantic representation functionality, Network 3 demonstrates weaker semantic representation capabilities, at least within the dataset employed in this study. Although some literature supporting embodied cognition has reported semantic activations within the sensorimotor network (de Zubicaray et al., 2013), the reliability of these findings remains contentious. Recent studies utilizing naturalistic paradigms have also found that language models do not significantly predict semantic activations in the sensorimotor network, whereas the association cortex plays a more stable and reliable role in semantic representation (Chen et al., 2024; Huth et al., 2016). Furthermore, our results presented in Section 3.3.1 indicate that Network 3 possesses weaker—or even nonexistent—semantic representation functions compared to other networks. Consequently, parcels within Network 3 were not reassigned.

2.3.2 Validation test

After conducting reliability tests, we proceeded with validation tests by analyzing the similarities between semantic network partitions obtained using different semantic models (i.e., semantic ratings and word embedding from GPT-2) and between semantic network partitions and brain networks obtained using traditional approaches (i.e., resting-state RSFC and modularity analysis). Our hypothesis is that the proposed method, which constructs brain semantic networks based on multidimensional semantic similarity, can more accurately capture the similarities and differences

in the semantic representational functions of brain regions, providing an advantage over traditional methods in partitioning semantic networks.

To test this hypothesis, we examined the similarities between semantic network partitions derived from different semantic models, as well as their alignment with brain networks generated using traditional methods. We expect that, if the proposed clustering method is effective, network partitions derived from the rating-based model should exhibit significantly higher similarity to those derived from GPT-2 than to the network partitions generated by traditional methods. Furthermore, by comparing the similarity and differences between the partitions of semantic networks based on these two semantic models, we can evaluate the influence of semantic model choice on network partitioning.

Constructing semantic network partitions using word embedding. We used GPT-2 (Radford et al., 2019), one of the most widely utilized computational language models, to extract word embeddings from text stimuli. GPT-2 is an autoregressive model trained to predict the next token based on preceding text, and it has demonstrated strong performance on a variety of downstream NLP tasks, such as semantically coherent text generation (Li et al., 2023, 2024). Furthermore, numerous studies have extensively validated the ability of GPT-2-extracted word embeddings (semantic representations) to interpret and predict neural activity. (Caucheteux et al., 2023; Goldstein et al., 2022; Schrimpf et al., 2021; Sun et al., 2020). For this study, we used the pre-trained GPT-2 medium model, which has 24 hidden layers, each with a hidden representation dimension of 1024, and is publicly available on Hugging Face[1]. According to previous research (Toneva and Wehbe, 2019; Wang et al., 2022a), the middle layers of deep language models are particularly effective at encoding various linguistic features. Therefore, we extract word embeddings from the 11th layer of the GPT-2 medium model.

In line with previous studies (Vulić et al., 2020; Y. Zhang et al., 2023a), we randomly sampled 1,000 sentences per target word from the Chinese Wikipedia corpus[2]. These sentences were then input into the model, and we extracted the contextualized word vectors for the target words from

[1] https://huggingface.co/uer/gpt2-medium-chinese-cluecorpussmall

[2] https://dumps.wikimedia.org/zhwiki/latest/zhwiki-latest-pages-articles.xml.bz2

the 11th layer. The final embedding for each target word was obtained by averaging the 1,000 contextualized vectors, resulting in a 1,024-dimensional word embedding. Previous research has shown that averaging contextualized word embeddings produces vectors that are either competitive with or outperform those generated by static distributional semantic models (DSMs) (Schrimpf et al., 2021; Sun et al., 2020), suggesting that these embeddings capture richer semantic information.

We chose this approach rather than directly inputting the target word into GPT-2 for word embedding extraction because GPT-2 is a context-dependent model, where the same word can have different representations in different contexts. Using multiple contextual sentences can capture the diversity of the target word in various contexts, resulting in a more comprehensive and accurate representation. It has been demonstrated that averaging contextual embeddings generates vectors that contain more semantic information than those obtained by directly inputting the target word (Bommasani et al., 2020; Chersoni et al., 2021). Moreover, this choice aligns with the hypothesis that context-independent conceptual representations are abstractions derived from token exemplar concepts (Yee and Thompson-Schill, 2016).

After extracting word embedding from GPT-2, we used the proposed method to construct semantic network partitions (detailed in Section 2.2).

Constructing brain network partitions using traditional methods. We chose two widely used traditional methods for constructing brain network partitions in neuroscience studies: the resting-state functional connectivity (RSFC) method and the modularity approach. For the resting-state RSFC method, we selected the widely used Yeo-7 network partition (Thomas Yeo et al., 2011) as the representative resting-state network. Thomas Yeo et al. (2011) provided heuristic labels for these seven networks, commonly referred to in the neuroimaging literature as Default, Somatosensory-Motor (SomMot), Dorsal Attention (DorsAttn), Salience/Ventral Attention (SalVentAttn), Limbic, Executive Control (Cont), and Visual (Vis). While we adopted these labels for descriptive convenience, as noted by Thomas Yeo et al., these names do not necessarily correspond precisely to the networks' functional roles.

For the modularity approach, following previous studies (Cao et al., 2014; Godwin et al., 2015; Moraschi et al., 2020; Rubinov and Sporns, 2011), we constructed cerebral cortex networks using

the commonly adopted modularity method on the same concept comprehension dataset. We applied Schaefer's parcellation, dividing the cerebral cortex into 1,000 regions of interest (ROIs), and constructed node-based connectivity matrices (1,000 × 1,000) for each subject. These matrices were then averaged to create a group-level connectivity matrix. We used the Louvain method (Blondel et al., 2008) to determine the optimal modularity partition, maximizing the quality function (Q) (Rubinov and Sporns, 2011) that reflects partition quality, as implemented in the Brain Connectivity Toolbox[3]. The optimal modularity partition was defined as the one with the highest Q value over 1,000 iterations.

Assessing the similarity between two semantic network partitions. After obtaining the corresponding brain networks, we compared the similarity between other brain network partitions and semantic brain networks by analyzing the distribution of each semantic network's voxels across networks defined by the other partitions. Taking the Yeo-7 network as an example, for each semantic network, we first identified the voxels within each semantic network's spatial mask and then quantified their overlap with the spatial masks of the Yeo-7 networks. The overlap between a semantic network and a particular Yeo-7 network was defined as the proportion of overlapping voxels relative to the total number of voxels in the semantic network. A higher proportion signifies greater similarity between the two networks. For example, if Semantic Network 1 contains N voxels in total, and X_{vis}, X_{def}, X_{lim}voxels overlap with the Visual, Default, and Limbic Yeo-7 networks respectively, the proportions are calculated as follows: X_{vis}/N, X_{def}/N and X_{lim}/N, representing the proportions of Semantic Network 1 associated with the Visual, Default, and Limbic networks, respectively.

Assessing potential biases introduced by resting-state parcellations. We utilized randomly parcellated parcels to construct semantic brain networks, thereby further demonstrating the stability of the proposed semantic networks. Specifically, we employed the region growing method (Adams and Bischof, 1994; Lu et al., 2003) for parcellation. Based on the predefined number of voxels per parcel, we randomly selected 783 seed points as initial region centers. From each seed point, we iteratively expanded into the unassigned voxels within their six-neighborhood, prioritizing voxels closest to the region center. To prevent overlapping regions and control region size, each voxel was assigned to only one region, and each region was limited to a

[3] http://www.brain-connectivity-toolbox.net

maximum number of voxels. Any remaining unassigned voxels were allocated to the nearest region, followed by overlap checks, thereby completing the cortical parcellation. The final parcellation divided the cortical surface into 783 parcels, each containing approximately 50 voxels. Subsequently, we applied our proposed method to map each parcel into the semantic space and performed clustering to obtain seven semantic brain networks.

Evaluating the stability of the semantic brain networks constructing using word embedding. We validated the semantic brain networks derived from the 11th layer of GPT-2 medium representations through three distinct approaches. First, we assessed whether different layers within GPT-2 medium could produce relatively stable semantic network partitioning results. Specifically, we constructed brain networks using representations from layers 10, 12, 13, 14, and 15 of GPT-2 medium and calculated the correlations between these network partitionings and the network partitioning obtained from layer 11. Second, we examined whether GPT-2 models with varying numbers of parameters could generate relatively stable semantic network partitions. We utilized representations from GPT-2 Distil (layer 5), GPT-2 Small (layer 8), and GPT-2 Large (layer 27) to construct brain networks and computed pairwise correlations with the network partitioning derived from GPT-2 medium. Third, we investigated whether word embeddings extracted from different corporas could lead to relatively stable semantic network partitions. For this analysis, we employed the Xinhua[4] and Chinese Wikipedia corpus respectively to extract word vectors and construct the corresponding brain network partitions.

2.4 Examining the cognitive functions of the partitioned semantic networks

In previous analyses, seven brain networks were delineated based on the multidimensional semantic-representation functions of various brain regions, and the reliability and validity of these partitions were evaluated. In this section, we investigated the cognitive functions associated with each of the seven semantic networks (See Figure 1d).

[4] http://www.xinhuanet.com/whxw.htm

2.4.1 Examining the multidimensional semantic-representation functions of the partitioned semantic networks

In this section, we evaluated each semantic network's capacity to represent semantic information across various dimensions. This capacity is reflected in the network's ability to decode semantic information pertaining to individual dimensions from its brain activations. Specifically, we evaluated how accurately each semantic network's activtations can predict 59-dimensional semantic ratings (detailed in Section 2.1.2). This method is usually used to reveal the internal information (or knowledge) encoded in text-based word vectors generated by different language models (Chersoni et al., 2021; Utsumi, 2020). For each semantic network, we mapped the t-value representations (detailed in Section 2.1.1) to each dimension of semantic ratings to obtain predicted semantic ratings using RSE (detailed in Section 2.5). Prediction performance was evaluated by calculating Pearson correlations between actual and predicted semantic rating vectors for each semantic feature. A high correlation for a particular semantic feature indicates that the network contains rich semantic information for that feature.

2.4.2 Tracing the source of semantic information acquisition in the partitioned semantic networks

Based on the dual-coding theory of semantic representation (Bi, 2021; Paivio, 1990), semantic information is acquired from two primary sources: perceptual experiences and linguistic experiences. To investigate the origins of semantic information acquisition, we assessed the decoding capabilities of specific brain networks' semantic activations using models of different modalities, including language models, visual models, and hybrid models. We hypothesize that the greater the similarity between the information acquisition sources of a model and those of a brain network, the more analogous their semantic representation functions will be.

Specifically, following the methodology outlined in Section 2.4.2, we mapped the t-value representation of each network and the hidden representation of different deep neural network models onto 59-dimensional semantic ratings. This enabled us to obtain the 59-dimensional semantic representation for each network and each deep neural model. We then calculated the Pearson correlation between the semantic representation of each network and each deep neural model to determine the semantic correlation coefficient.

To ensure the robustness of our analysis, we selected a diverse range of deep neural network models, representing major architectural types across each modality: language (e.g., BERT-family, GPT-2 (Zhao et al., 2019), T5 (Zhang et al., 2021)), visual (e.g., BEiT (Bao et al., 2022), ViT (Dosovitskiy, 2020), DEiT (Touvron et al., 2021)), and multi-modal (e.g., CLIP (J. Zhang et al., 2023), OFA (P. Wang et al., 2022), ViLT (Kim et al., 2021)). For each architectural type, we selected models trained on various datasets using different methods. For instance, the BERT-family includes BERT (Devlin, 2018), MacBERT (Cui et al., 2021), RoBERTa (Cui et al., 2020), ERNIE (Sun et al., 2021) and Roformer (Su et al., 2024). These models build upon BERT by introducing various enhancements and modifications to optimize performance for specific tasks or scenarios. We did not finetune any of these deep neural network models ourselves but leveraged the pretrained models available on Huggingface[5]. Details of the specific pretrained model checkpoints are described in supplementary material. For each semantic network, we averaged the semantic correlation coefficients of each model within the same architectural type to obtain the final semantic correlation coefficient. A high correlation indicates a strong semantic resemblance between the brain's semantic network and the deep neural model, suggesting that this semantic network acquires information from the same modality as the deep neural model.

For deep neural network models across different modalities, we employed various methods to extract the hidden layer representations for each target word (Chersoni et al., 2021; Y. Zhang et al., 2023a). For deep neural language models, we randomly sampled 1,000 sentences for each target word from the Chinese Wikipedia corpus. We then fed these sentences to the models and extracted the vectors from each layer. The final embedding for each target word was obtained by averaging the 1,000 contextualized vectors, resulting in a 1,024-dimensional word embedding. For deep neural visual models, S. Wang et al. (2022b) published six different images depicting each word. We fed these six images to the models and extracted the vectors from each layer. The image representations were obtained by averaging the six vectors from the images of the target word. For deep neural multi-modal models, we provided both the target word and its six corresponding images as inputs and extracted vectors from each layer of the models. Multi-modal representations were obtained by averaging the six image vectors together with the text vector of the target word. Since ViLT and OFA currently lack open-source Chinese pre-trained models, we translated the target word into English and paired it with six different images. The vectors from each layer were then extracted from both ViLT and OFA.

[5] https://huggingface.co/

2.4.3 Examining the involvements of the partitioned semantic networks in general domains of cognitive functions

Semantic representation functions are intricately connected to numerous cognitive domains, resulting in brain regions responsible for semantic processing frequently participating in a wide array of cognitive activities. In this section, we leveraged the decoding capabilities of the large-scale functional neuroimaging database Neurosynth (Yarkoni et al., 2011) to investigate the involvement of partitioned semantic networks in various cognitive functions. This approach enables us to further elucidate the distinct characteristics of different semantic networks in their roles within cognitive processes.

The Neurosynth Image Decoder [6] associates each of 1,335 keywords with a unique meta-analytic map. These keywords encompass brain structures such as the prefrontal cortex, hippocampus, amygdala, and thalamus, as well as cognitive functions like working memory, attention, language, and emotion. In our study, we inputted each subnetwork of the semantic network partition into the decoder. It then calculated Pearson correlations between the semantic subnetwork and the meta-analytic maps associated with each keyword, subsequently ranking the keywords in descending order of their correlation coefficients.

From the ranked list, we retained only keywords relevant to cognition. To reduce redundancy, singular and plural forms of the same word (e.g., "language" and "languages") were consolidated. Finally, for each subnetwork, the top 10 functionally relevant keywords, based on their correlation values, were selected to construct word clouds.

Furthermore, we utilized the Yeo-7 network to examine the involvement of partitioned semantic networks within general domains of cognitive functions. The Yeo-7 network is characterized by well-defined cognitive functions. For instance, the Default Mode Network is associated with higher-order cognitive processes such as introspective thinking, self-referential processing, memory retrieval, and future planning. The visual network plays a critical role in both primary and higher-level visual processing, including visual perception, image recognition, spatial navigation, and the regulation of visual attention. Therefore, employing the methodology outlined

[6] http://www.neurosynth.org/decode/

in Section 2.3.2, which involves analyzing the distribution of each semantic network's voxels across the networks defined by the Yeo-7 networks, we can further infer the involvement of partitioned semantic networks in broader cognitive function domains.

2.5 Representational similarity encoding

To map representations E derived from semantic networks or deep neural network models onto 59-dimensional semantic ratings C, we utilized the representational similarity encoding method (Anderson et al., 2016). This approach, referred to as RSE, calculating the similarities between different word representations within E. Subsequently, it reconstructs the semantic rating of each word by computing a weighted average of the semantic ratings of other words, where the weights are determined by the previously calculated similarities. The procedures of SEA are outlined below:

Initially, a set of word representations $E = \{e_1, e_2, \ldots, e_n\}$ is extracted from various deep neural network models or semantic networks. For each pair of representations (e_i, e_j), their similarity is quantified using the Pearson correlation coefficient (ρ). Consequently, a similarity matrix M is formed, where $M \in \mathbb{R}^{n\times n}$ and $M_{ij} = \rho(e_i, e_j)$.

Subsequently, for semantic networks or deep neural network models, we posit that if network-based or model-based representations encode the same information as specific semantic feature vector, the similarity relation of the network-based or model-based vectors and that of the semantic feature vector is the same. We can predict each semantic vector by multiplying the above similarity matrix with corresponding semantic vector. To mitigate the influence of actual semantic values, we subtract the real semantic feature vectors from the predicted vectors to derive the predicted semantic feature matrix:

$$C' = (M - I_n)C \tag{7}$$

where $C \in \mathbb{R}^{n\times m}$ represents the real semantic feature vectors, $C' \in \mathbb{R}^{n\times m}$ denotes the predicted semantic feature vectors, and $I_n \in \mathbb{R}^{n\times n}$ represents a square matrix in which all the elements of the main diagonal are 1, and all other elements are 0.

Finally, the Pearson correlation is employed to assess the similarity between the predicted and the actual semantic feature vectors in each dimension:

$$r = (\rho(C_{(:,1)}, C'_{(:,1)}), \rho(C_{(:,2)}, C'_{(:,2)}), \cdots, \rho(C_{(:,n)}, C'_{(:,n)})) \tag{8}$$

A higher correlation score for a given dimension of r indicates that the corresponding semantic network or deep neural network model captures a broader range of information related to that semantic dimension.

3 Results

3.1 Semantic network partition

The semantic network partition drived from multi-dimensional semantic ratings are presented in Figure 2. These semantic networks exhibit bilateral symmetry across the hemispheres. Supplementary Figure S2 provides a comprehensive explanation of the cluster selection process, specifically outlining the determination of the optimal k value. These figures show an elbow point occurring between k=6 and k=9, with k=7 identified as the optimal choice based on the silhouette coefficient and cross-subject stability. Overall, our approach successfully delineated seven distinct semantic networks. Among them, five brain networks demonstrate strong correspondences with the neural associations of specific semantic-related cognitive functions reported in the literature.

Networks 1 and 4 are primarily localized in the ventral, dorsal temporal and occipital cortices. These regions are associated with object visual recognition and are believed to be involved in the representation of visual object semantics (Ludersdorfer et al., 2019; Martin, 2007). **Network 5** is mainly situated in the lateral frontal and temporal cortices, corresponding to classical language areas such as Broca's and Wernicke's areas. This network is thought to participate in the representation of semantic information based on linguistic symbol encoding (Bi, 2021). **Network 6** predominantly includes the medial temporal lobe and the ventral and lateral parietal-occipital cortices. These regions are related to spatial navigation and are considered to be involved in spatial semantic encoding (Epstein, 2008; Epstein et al., 2017; Lin et al., 2024). **Network 7** has a widespread distribution. Visual inspection indicates significant overlap with the social brain network, including the dorsolateral prefrontal cortex, temporoparietal junction, anterior superior temporal sulcus, and posterior cingulate cortex. These regions are believed to be involved in the representation of social semantics (Lin et al., 2020, 2019, 2018b; Patel et al., 2021). The remaining two networks, **Networks 2 and 3**, exhibit very extensive distributions across various brain regions. Although parts of these networks have been identified in the literature as participating in semantic representation—for instance, the motor regions within Network 3 have been implicated in motor semantics (Carota et al., 2017; Dreyer and Pulvermüller, 2018;

Fernandino and Iacoboni, 2010), and the anterior temporal lobe (ATL) within Network 2 is considered a semantic hub (Holland and Lambon Ralph, 2010; Zhao et al., 2017)—it remains challenging to comprehensively associate these networks with specific types of semantic representations documented in existing studies.

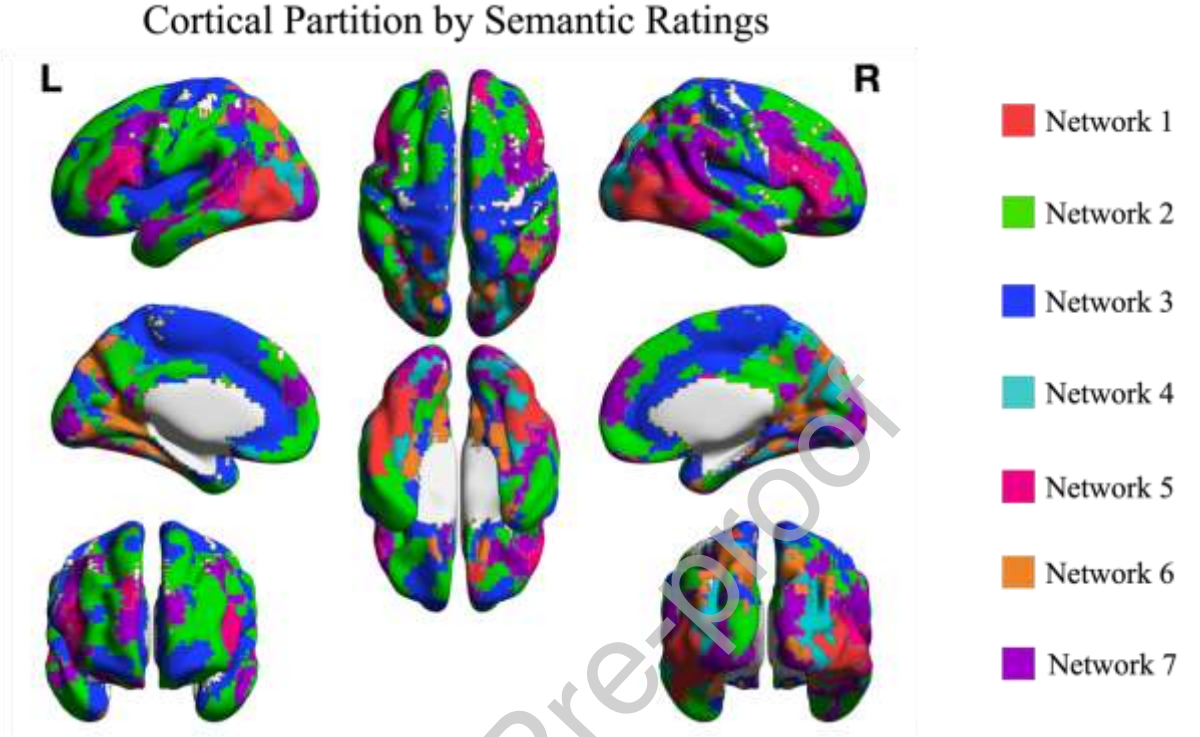

Fig 2. Cortical semantic network partition. The cerebral cortex, mapped by multi-dimension semantic ratings, was divided into seven principal semantic networks.

3.2 Assessing quality of semantic network partition

3.2.1 Reliability test

We first assessed the reliability of our network partition at the global level. Figure 3a illustrates the high consistency of semantic networks derived across various subject sampling subsets, clustering algorithms, and cortical parcellation methods. As shown Figure 3a (left), regardless of which half of the subjects are selected, the correlation between the obtained network partitions is consistently higher than 0.6 ($p < 0.000001$), with the highest correlation reaching 0.76. This result indicates that the proposed network partitioning method is not sensitive to variations in subject selection. Figure 3a (middle) demonstrates that the correlation between network partitions obtained by different clustering methods is significantly higher than 0.6 ($p < 0.000001$), with the highest correlation reaching 0.82. This suggests that the semantic representation matrix has a relatively clear clustering structure, allowing different clustering algorithms to capture similar patterns. This further illustrates that the semantic representation of each parcel (detailed in

Section 2.1.1) effectively reflects the differences in semantic information between parcels, with those in similar brain networks exhibiting more similar semantic representations. Furthermore, the higher correlation between network partitions created by K-means and GMM may be attributed to their similarity in distance measures (Patel and Kushwaha, 2020). Figure 3a (right) shows that the correlation between network partitions created by different cortical parcellations is also higher than 0.6 ($p < 0.000001$), with over 80% of cortical parcellations exhibiting correlations above 0.7. This suggests that our method can effectively extract the semantic information contained in each parcel, regardless of whether the cerebral cortex parcellation is divided according to structural pattern or resting-state function, and regardless of the resolution of the parcellation, and then construct a relatively consistent semantic network partition. However, the semantic network results derived from the BN atlas, which comprises 210 cortical parcels, exhibited slightly lower similarity compared to those obtained using higher-resolution cortical parcellations. This discrepancy may be attributable to the factors discussed in Method 2.2, wherein utilizing a lower-resolution atlas increases the likelihood that voxels with highly heterogeneous semantic representation functions are aggregated within the same parcel.

We further evaluated the semantic network partition quantitatively at the local level by assessing two metrics for each parcel and network: network assignment confidence and inter-subject connectivity consistency (referenced in Figure 3b and c). We observe that the overall network assignment confidence scores are relatively low, which may be attributed to three factors. First, the differences in semantic information distribution between networks are not particularly large (as shown in Figure 5). Second, the calculation method is highly stringent. Previous studies that used functional connectivity of parcels as representations also obtained relatively low network assignment confidence scores (Ji et al., 2019). Third, the exceptionally low confidence scores for Network 2 and Network 3 have significantly lowered the overall average.

The lower confidence score of Network 2 (and similarly Network 3) may be attributed to three potential factors: a low signal-to-noise ratio (SNR), low inter-subject semantic consistency, and limited semantic information functionality. First, Network 2 has a lower SNR of 42, compared to an average of 46 across other networks. A Pearson correlation analysis confirmed a significant relationship between SNR and confidence scores ($r = 0.20$, $p < 0.00001$), indicating that lower SNR increases noise, thereby reducing confidence scores.

Second, lower inter-subject semantic consistency may introduce noise by averaging distinct semantic representations across different brain parcels, which could reduce confidence scores. To assess this, we calculated the mean similarity of cortex-wide semantic representations for each parcel across subjects. A higher inter-subject value indicates greater semantic consistency among subjects, whereas a lower value reflects greater semantic variability. As shown in Figure 3c, the inter-subject semantic consistency was high, relative to the theoretical minimum of 0 (which represents minimum consistency). Network 1 demonstrated the highest semantic consistency (the highest inter-subject value), corresponding to its higher confidence score. In contrast, Network 2 and Network 3 showed lower semantic consistency (the lowest inter-subject value), consistent with its lower confidence score.

Third, Network 2 and Network 3 contains weaker semantic information functions than other networks. As shown in Figure 5, semantic networks are ranked from top to bottom based on the richness of their semantic information, quantified by the average Pearson correlation coefficient across all semantic dimensions. The ordering of semantic information richness across networks strongly correlates with the ordering of mean assignment confidence scores. As seen in Figure 3c, Networks 5 and 1, which have high semantic information functions, also have high network assignment confidence scores. Conversely, Network 2 and 3, with low semantic information functions, has low network assignment confidence scores. These findings indicate that networks with higher semantic representation functions tend to exhibit greater stability in brain network partitioning.

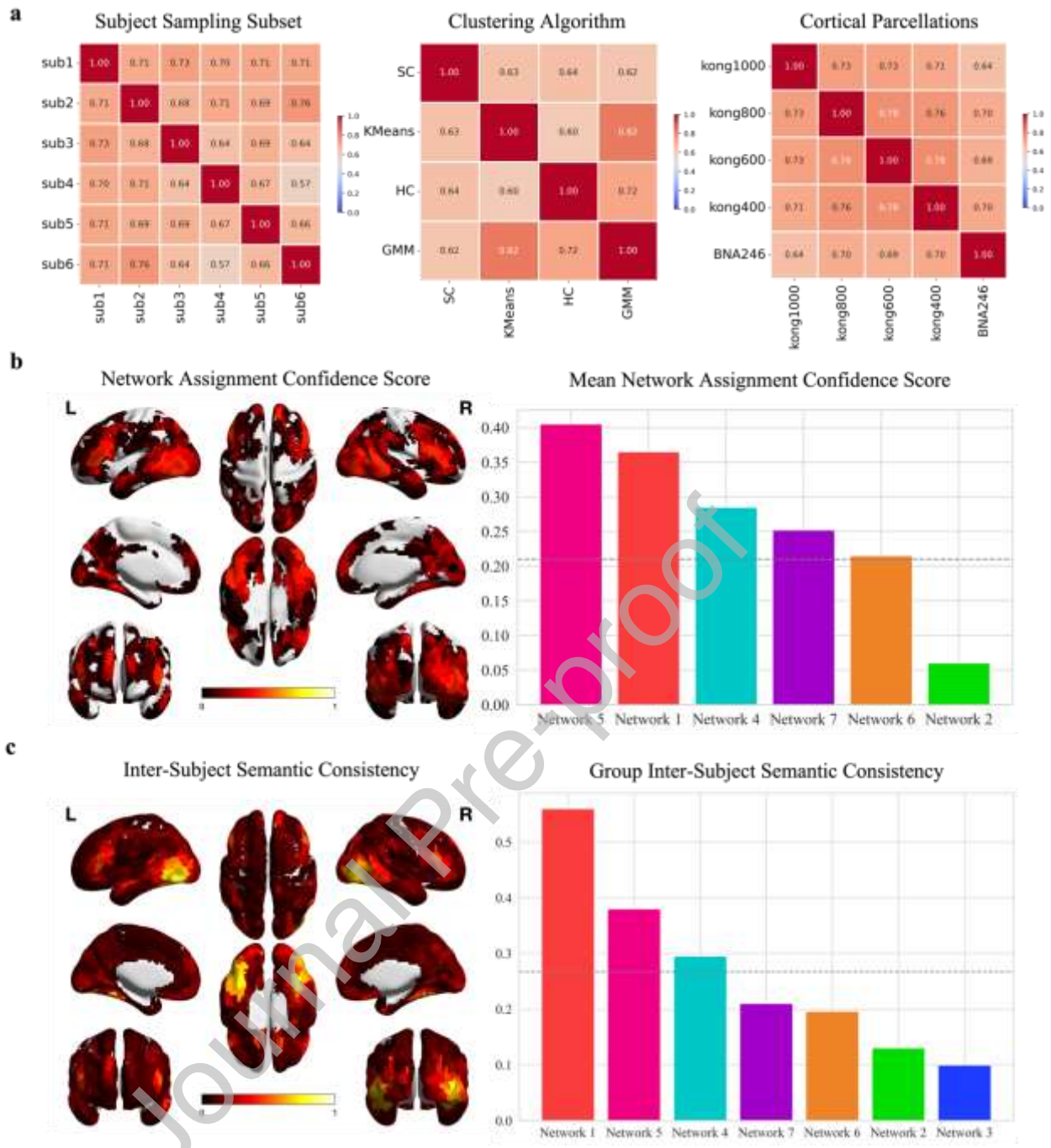

Fig 3. Reliability test of cortical semantic network partition. A) Pearson correlation of cortical semantic network partitions obtained by utilizing different half-split subjects, different clustering algorithms, and distinct cortical parcellations. B) Left: The cortical map featuring network assignment confidence scores, indicating the semantic similarity of each parcel to its assigned network. Network 3 has a negative confidence score, which is not shown on the figure. Right: Network-level averages of parcel-level confidence scores (including all seven networks). C) Left: The cortical map exhibiting inter-subject semantic consistency, quantifying similarity in semantic distribution patterns across subjects for each cortical parcel. Right: Network-level averages of parcel-level inter-subject semantic consistency.

3.2.2 Validation test

Figure 4 shows that the similarities between semantic-network partitions obtained using different semantic models (i.e., interpretable semantic ratings and GPT-2), and between semantic-network partitions and brain networks obtained using traditional approaches (i.e., resting-state RSFC and modularity analysis).

We observe that, compared to the similarity between networks defined by the Yeo-7 network and those derived from modularity-based mapping, the similarity between networks partitioned using two different semantic models is significantly greater. Figure 4a (middle) illustrates that most networks exhibit a clear one-to-one correspondence between the partitions derived from the two semantic models, such as Network 1, Network 4, Network 5, Network 6, and Network 7. This indicates that these networks maintain strong consistency across different semantic models, further suggesting that our method demonstrates high cross-model stability. However, Networks 2 and 3 from the semantic networks partitioned using semantic ratings exhibit the greatest similarity to Network 2 from the semantic networks partitioned using GPT-2. These two networks were previously identified as containing less semantic functions. This suggests that parcels with weaker semantic representations are more susceptible to model-specific influences in our approach, an issue that we discuss in greater detail in the Discussion section. Figure 4b (middle) shows that several networks from the semantic-network partitions (i.e., Network 1, Network 4, Network 6) exhibit the greatest similarity with Network 1 within the Yeo-7 network. Networks 2 and 7 within the semantic-network partitions are most similar to Network 7 in the Yeo-7 network. Similarly, Figure 4c (middle) reveals that Networks 1, 4, and 6 from the semantic-network partitions are most similar to Network 2 derived from modularity analysis, while Networks 2 and 7 from the same partitions are most similar to Network 4 from modularity analysis. These findings suggest that brain networks obtained using traditional approaches (i.e., Yeo-7 and modularity-based networks) exhibit low similarity with the networks derived from our semantic partitioning method. In summary, the considerable similarity between network partitions derived from two highly distinct semantic models suggests that the proposed method effectively captures the brain network organization patterns specific to semantic representational functions and demonstrates robustness across varying semantic models.

It can be noticed from Figure 4c (right) that the modularity method divides the brain into 12 networks, eight of these modules are small, the largest of these eight networks accounts for only

0.28% of the cortical voxels. These small modules lack clear biological functional interpretations. Additionally, the four main networks identified through modularity closely resemble the hierarchical structure observed in resting-state data (Margulies et al., 2016), effectively separating primary sensorimotor and transmodal regions, as well as distinct regions within the primary sensorimotor areas (e.g., somatomotor, auditory, and visual cortex). However, modularity-based partitioning fails to capture the task-related semantic functions that influence network division. This limitation likely arises because modularity analysis primarily relies on the network's topological structure and is not designed to disentangle multidimensional, semantic-related information from fMRI data. As a result, it partitions the cortical network based solely on the topological features of brain regions, rather than accounting for the semantic representation functions associated with those regions.

Supplementary Figure S4 represents the cortical semantic network partition, consisting of seven major networks, derived from Schaefer's parcellation and the random parcellation. Visual inspection reveals that the spatial patterns of the two semantic network partitions are largely similar, particularly for networks with high semantic functions. The majority of networks (Networks 1, 2, 5, 6, and 7) exhibit substantial spatial overlap. This consistency underscores the stability of our derived parcellation approach.

As illustrated in Supplementary Figures S6 and S7, the brain network partitions obtained from different layers and models exhibited high correlations, indicating consistent clustering across various GPT-2 configurations. Additionally, Supplementary Figure S8 demonstrates that brain network partitions derived from different corporas showed substantial overlap in specific brain regions with high semantic functions, such as Networks 1, 4, 5, and 6. These findings collectively suggest that the proposed clustering method yields relatively stable semantic network results across different layers, model sizes, and corpora.

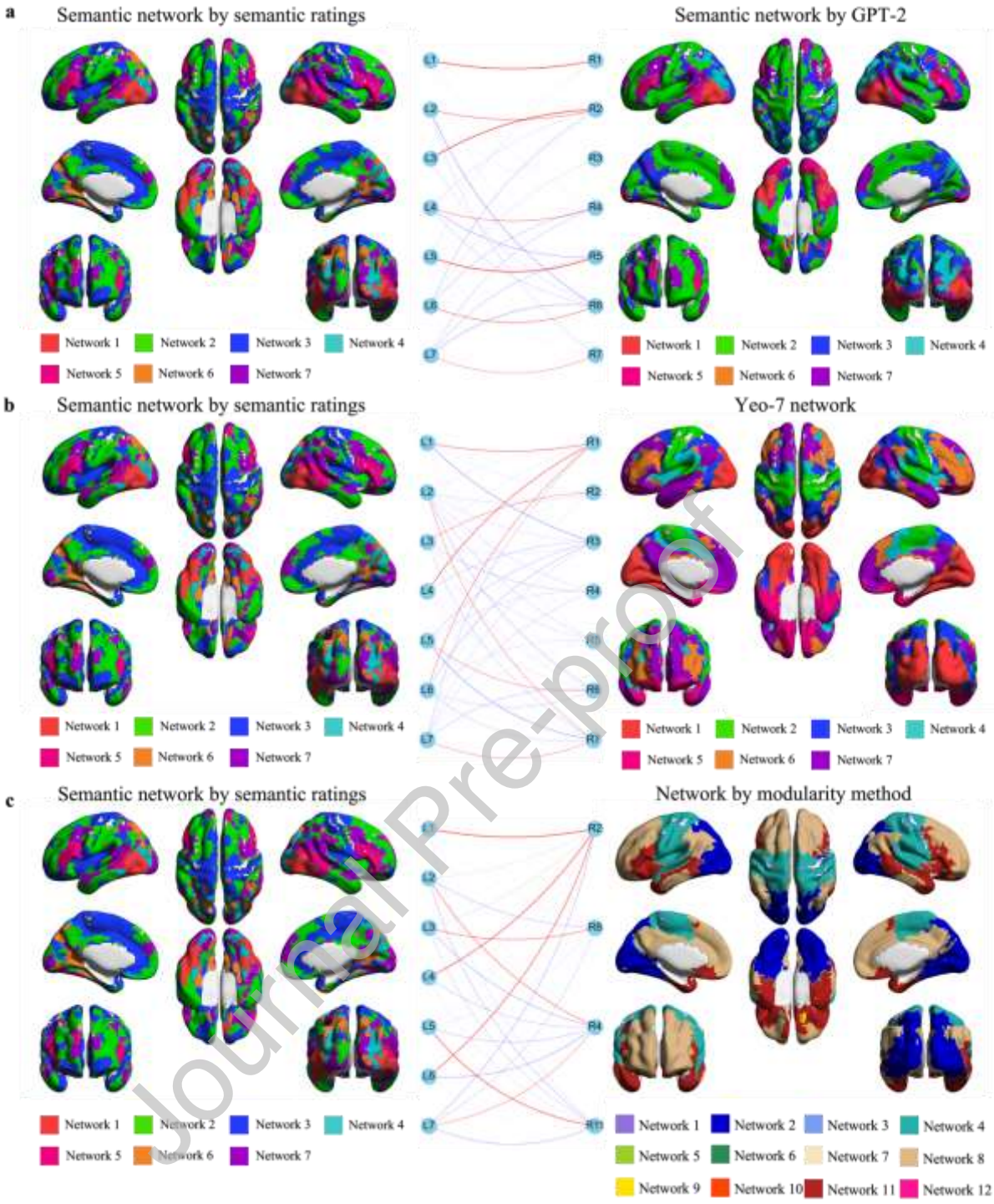

Fig 4. Validation test of cortical semantic network partition. A) Semantic networks mapped by semantic ratings (left) and word embeddings extracted from GPT-2 (right), was divided into seven principal semantic networks. Middle: The similarity between the left and right cerebral networks. Network Lx refers to the x-th semantic network mapped by semantic ratings (left), while Network Rx denotes the x-th semantic network mapped by GPT-2 (right). Connections between Network Lx and Network Rx indicate the proportion of voxels in Network Lx that are also present in Network Rx. Thicker connections signify a greater proportion of Network Rx

within Network Lx, thereby suggesting higher similarity between the corresponding brain networks. For example, the connection between Network L1 and Network R1 denotes the proportion of voxels in Network L1 that belong to Network R1. A red connection highlights that Network R1 has the highest voxel overlap with Network L1, indicating that these two networks exhibit the highest degree of similarity. B) Left: Semantic networks mapped by semantic ratings. Middle: The similarity between the left and right cerebral networks. Right: Resting-state network (Yoe-7 network). C) Left: Semantic networks mapped by semantic ratings. Middle: The similarity between the left and right cerebral networks. Right: Cortical networks, mapped using the modularity method, was divided into twelve networks. In the analysis of brain network similarity, we focused exclusively on four principal networks: Networks 2, 4, 8, and 11. These four networks collectively encompass 98.44% of the cortical voxels across the entire cerebral cortex. In contrast, the largest of the remaining eight networks accounts for only 0.28% of the cortical voxels.

3.3 Cognitive functions of the partitioned semantic networks

3.3.1 The multidimensional semantic-representation functions of partitioned semantic networks

Figure 5 displays the multidimensional semantic-representation functions of the partitioned semantic networks. Concerning sensorimotor dimensions, sensory information is encoded in the brain more than other motor information. For instance, the correlation of face and body dimensions is significantly higher than that of other motor dimensions, such as heavy and motion, which is potentially relevant to humans' ability to rapidly detect and recognize faces. In terms of non-sensorimotor dimensions, most spatial (e.g., near, toward, away, and path) and temporal (e.g., number, time, and duration) dimensions are significantly lower among nearly all attributes. Conversely, human, self, and cognition dimensions in social and cognitive domains achieve higher correlations. Additionally, we observe that some negative emotions (e.g., sad, angry, and disgusted) are better predicted by some networks than positive ones.

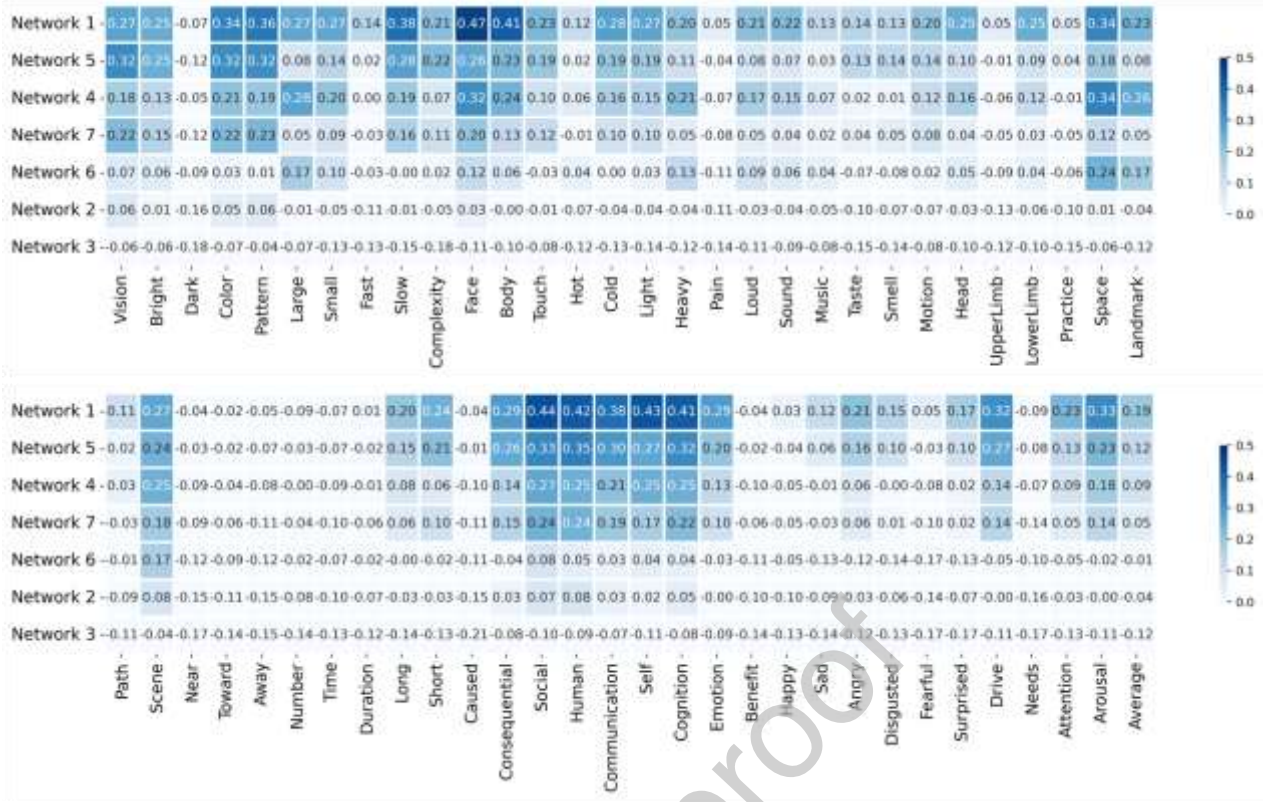

Fig 5. Multidimensional semantic-representation functions of the partitioned semantic networks. Pearson correlations between the estimated semantic features and the original data for each network. Semantic networks are ranked from top to bottom based on the richness of their semantic information, which is quantified by the average Pearson correlation coefficient across all semantic dimensions. A high correlation for a particular semantic feature indicates that the network contains rich semantic information for that feature.

3.3.2 The source of semantic information acquisition in the partitioned semantic network

The premise of using deep neural network models to explore the information acquisition of semantic networks is based on the observation that different deep neural network models exhibit distinct distributions of semantic information. As illustrated in Figure 6a, models within the same modality demonstrate consistent semantic information distribution, whereas models across different modalities displayed significant variation. In sensorimotor domains, language models show a higher correlation in the visual domain and performed comparably in other domains such as somatic, audition, gustation, and olfaction. Visual models, on the other hand, exhibit similar performance in visual, somatic, gustation, and olfaction domains but show relatively low correlation in the audition domain. In non-sensorimotor domains, language models perform comparable in social, cognition, and emotion domains, while visual models had low correlation in

the emotion domain. This finding supports prior research indicating that text information significantly contributes to the affective content of lexical items (Lenci et al., 2018; Recchia and Louwerse, 2015). The information distribution of multi-modal models integrates the patterns observed in both language and visual models. These results suggest that models from different modalities exhibit distinct patterns of semantic information distribution.

Figure 6b displays the correlation between network partitions and deep neural network models. Both language and visual models show high correlation in areas related to Network 1, Network 2, Network 5, and Network 7. However, language models exhibit lower correlations in Network 4 and Network 6. These findings imply that networks like Network 1, Network 2, Network 5, and Network 7 obtain semantic knowledge through both language and visual experiences, whereas Network 4 and Network 6 primarily rely on visual inputs. Additionally, multi-modal models, particularly single-stream models such as ViLT and OFA, demonstrate the highest correlations across most networks. Yet, in Network 4 and Network 6, visual models show correlations comparable to those of multi-modal models, also indicating that these networks may rely less on language experiences for semantic knowledge acquisition.

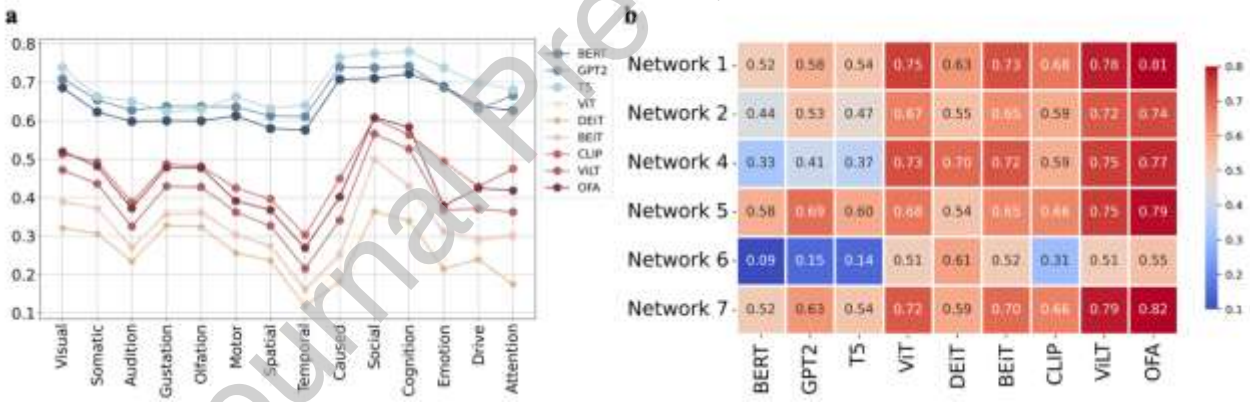

Fig 6. The source of semantic information acquisition in the partitioned semantic networks. A) Pearson correlations per domain between the estimated semantic features and the original data for each model. B) Semantic correlation between each semantic network and neural model indicated their similarity in semantic information. A high correlation indicates a strong semantic resemblance between the brain's semantic network and the deep neural model, suggesting that this semantic network acquires information from the same modality as the deep neural model.

3.3.3 The involvements of the partitioned semantic networks in general domains of cognitive functions

Figure 7 shows the involvements of the partitioned semantic networks in general domains of cognitive functions. It can be noticed from Figure 7a that voxels within each semantic network intersect with multiple resting-state networks defined in the Yeo-7 framework. For instance, the voxels in Network 1 predominantly overlap with the visual (53.1%) and dorsal attention (34.4%) networks in the Yeo-7 framework. Similarly, voxels in Network 5 primarily overlap with the default (35.8%) and executive control (33.6%) networks. This finding suggests that the brain's internal network patterns undergo dynamic reorganization to varying degrees during task execution compared to the resting state. The subsequent paragraphs will individually examine the involvement of each partitioned semantic network within the general domains of cognitive functions, drawing on the findings presented in Figures 5 and 7.

Network 1 is highly specialized in visual processing, with a focus on recognizing faces, objects, and other visual stimuli. The significant presence of voxels in the Visual and DorsAttn networks underscores its critical role in visual perception and attention. Key functional terms for this network include objects, visual, face, and motion viewing. Additionally, Network 1 contains semantic information related to faces, social cues, self-referential processing, and body recognition. This suggests its involvement in both visual element recognition and social perception, contributing to self-awareness within visual contexts.

Network 2 is involved in higher-order cognitive functions, encompassing default mode processing, tactile perception, referential thinking, and theory of mind. The predominance of voxels in the Default and Cont networks highlights its role in introspection, social cognition, and executive functions. Key functional terms include default, tactile, and theory of mind. The semantic information associated with human interactions, social dynamics, and cognitive activities emphasizes Network 2's importance in understanding social behavior and mental states.

Network 3 is primarily associated with somatosensory and motor functions. The high concentration of voxels in the SomMot and SalVenAttn networks underscores its role in processing sensory and motor information. Key functional terms include somatosensory, motor, and pain. The absence of semantic information suggests that Network 3 is specialized for direct sensorimotor tasks rather than higher-order semantic processing.

Network 4 is dedicated to visual and spatial processing, with a strong emphasis on object and face recognition, spatial awareness, and perceptual attention. The predominance of voxels in the Visual and DorsAttn networks highlights its role in visual perception and spatial orientation. Key functional terms include visual, objects, and spatial. The semantic information, including space, face, body, and landmarks, indicates Network 4's role in integrating visual stimuli with spatial and bodily awareness, essential for navigating and interacting with the environment.

Network 5 is specialized in language processing, including syntax, semantics, and phonology. Its voxels are mainly found in the Default and Cont networks, indicating its role in complex cognitive functions such as language comprehension and executive control. Key functional terms include sentence, language, and semantic. The semantic information related to humans, cognition, social interactions, and patterns suggests that Network 5 is crucial for understanding and producing language, as well as processing social and cognitive patterns.

Network 6 is involved in navigation, episodic memory, and spatial processing. The distribution of voxels across the Visual, DorsAttn, and Default networks indicates its role in integrating visual information with memory and attentional processes. Key functional terms include navigation, episodic memory, and place. The semantic information, including space, landmarks, scenes, and large objects, highlights Network 6's function in spatial navigation and memory retrieval, essential for orienting oneself in the environment.

Network 7 is primarily associated with theory of mind, mentalizing, and social cognition. The presence of voxels in the Default, Visual, and Cont networks suggests its role in understanding others' mental states, beliefs, and intentions. Key functional terms include theory of mind, beliefs, and mental states. The semantic information related to social interactions, human cognition, vision, color, and patterns indicates that this network integrates cognitive and visual information to facilitate social understanding and interpersonal communication.

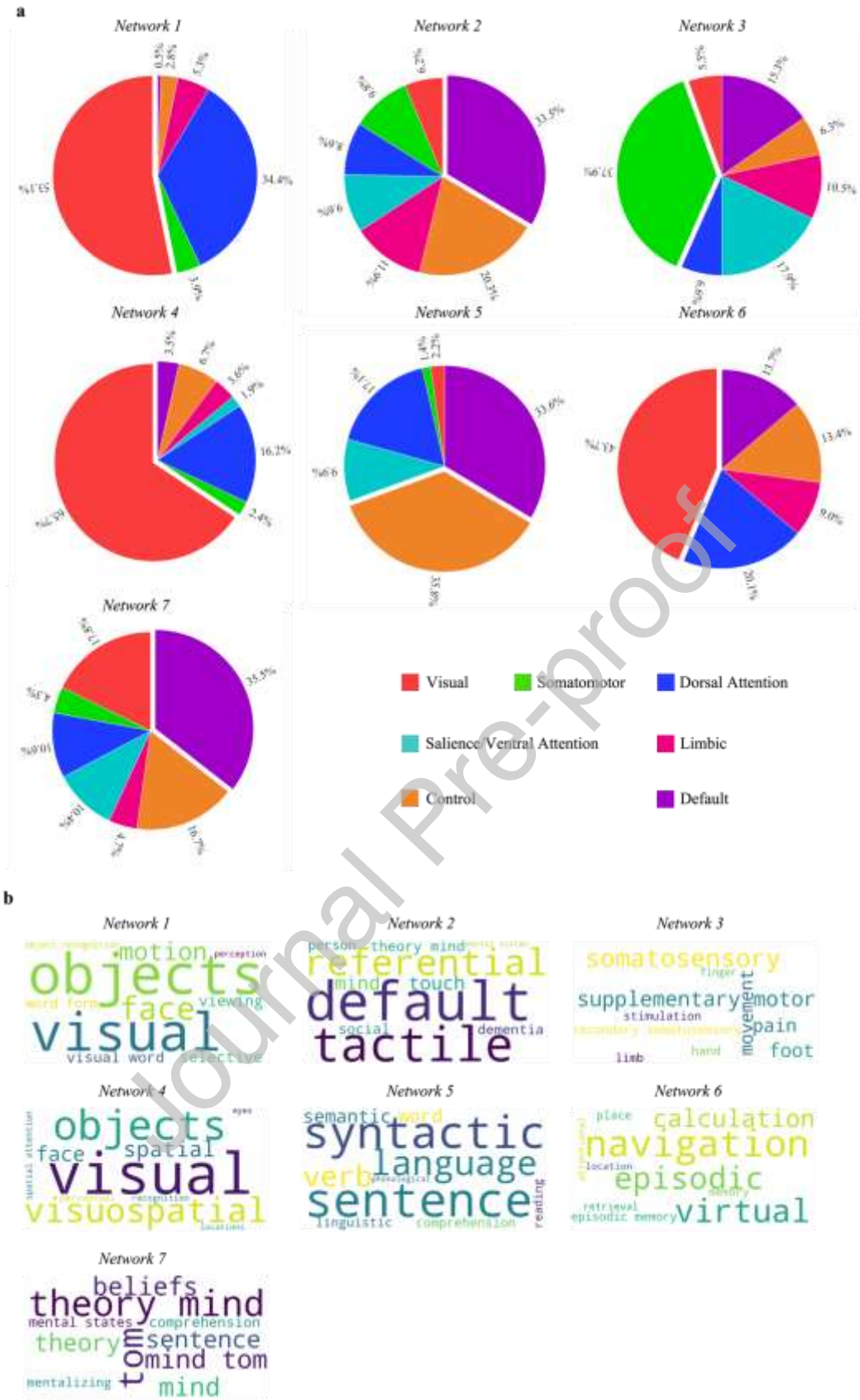

Fig 7. Involvements of the partitioned semantic networks in general domains of cognitive functions. A) The proportion of voxels within each semantic network that belong to each of the Yeo-7 networks. B) Key functional terms associated with each cortical semantic network.

4 Discussion

We propose a novel method for constructing large-scale brain networks based on specific cognitive functions and evaluate it using fMRI data from a concept comprehension task. We demonstrate the method's reliability and cross-semantic model stability, revealing significant differences between network partitions obtained using the proposed method and traditional brain network partitions (such as those based on resting-state RSFC and modularity analysis). Further analysis indicates that the different semantic networks we partitioned exhibit systematic differences in terms of multidimensional semantic representation, sources of semantic acquisition, and their associations with general cognitive domains. Additionally, the strength of semantic representation functions is correlated with the stability of brain regions within the semantic network partitions.

We find that the cortical semantic network partition aligns with prior work on semantic representation to a certain extent. Network 1 and Network 4 show a high degree of overlap with the regions of significant activation identified in several fMRI studies investigating concept comprehension using image stimuli (Connolly et al., 2012; Devereux et al., 2013; Fu et al., 2023). Both extend from the lateral and medial fusiform areas to the lateral occipital cortex, the ventrolateral inferior temporal gyrus, the dorsolateral middle temporal gyrus, and the inferior parietal lobule. These regions encode high-order visual and semantic structures. Additionally, multiple fMRI studies on pure language tasks (using word stimuli and natural discourse stimuli) have found that the inferior frontal gyrus, middle frontal gyrus, and posterior superior temporal sulcus, which are sensitive to various semantic dimensions (i.e., motor, vision, space, social) (Fernandino et al., 2016; Lin et al., 2024), significantly overlap with Network 5.

In addition, we find that multiple brain networks converge near the left angular gyrus, while no similar phenomenon is found in the right angular gyrus. These findings suggest that the left angular gyrus contains diverse and rich semantic information, consistent with previous studies (Lin et al., 2024). These studies have found that neural correlates of different semantic dimensions are primarily concentrated near the angular gyrus. According to the research by Xu et al. (2016) , the angular gyrus is located at the intersection of the multimodal experiential semantic system and the language support semantic system, serving as a critical hub that facilitates communication between these two systems. Therefore, the present study provides further evidence that the angular gyrus may serve as a central region for semantic representation.

Furthermore, we find that almost all semantic networks represent a variety of semantic information, not just one kind (Figure 5). Most networks can effectively represent social-related semantic dimensions (i.e., social, human, communication, self, cognition). The networks that contain more social-related information also contain relatively more sensorimotor-related information. These findings support the perspective of Binder et al. (2009) that the general semantic network consists of convergence zones representing multi-modal sensory-motor semantics. Binder and Desai proposed that nearly all parts of the general semantic network are involved in social cognition, and the engagement of these areas in both social and nonsocial tasks reflect a common process: the retrieval of conceptual knowledge abstracted from sensory-motor experience. Supporting this, Fernandino et al. (2016) found that certain core regions of the general semantic network are sensitive to all major types of sensory-motor semantic attributes. The results of the current study, combined with those of Fernandino et al., provide consistent evidence supporting the theory proposed by Binder and Desai

Moreover, we find that several non-sensorimotor semantic dimensions, including near, toward, away, number, time, benefit, and needs, are not represented by most semantic networks. We first rule out the possibility that the stimulus words lacked this semantic information. Statistical analysis revealed that a significant proportion of words with high scores in these dimensions are present, with 52% of words scoring above 3 in the benefit dimension and 32% in the needs dimension. Several potential reasons for this absence are considered. The first reason could be that word-level stimuli are insufficient to evoke certain semantic dimensions in subjects. Lin et al. (2024) found that semantic-dimension effects were more pronounced in widespread brain areas during natural narrative listening compared to word comprehension tasks. This suggests that contextual semantic information in natural and sequential language processes plays a significant role. Reviews on affective neuroimaging using naturalistic stimuli, such as movies and stories, highlight that these complex stimuli elicit strong, multidimensional brain responses (Saarimäki, 2021). The second reason could be that these semantic dimensions lack neural reality. Although this cannot be determined definitively at present, future research can employ the proposed method to screen the semantic dimensions suggested by Binder et al. (2016) to further explore the fundamental semantic dimensions in the human brain. The third reason could be the prolonged scanning sessions, which may have caused negative emotions in the subjects, making it difficult for them to comprehend the meanings of certain stimulus words.

In this paper, we present two significant methodological contributions. First, we introduce a novel approach for constructing large-scale brain networks based on specific cognitive functions. The proposed method is generalizable to other task domains, provided that the target cognitive function can be decomposed into multiple dimensions, each with neural representational validity. Under these conditions, the method allows for the extraction of relevant dimensions from brain signals derived from parcels and the subsequent clustering of these dimensions to construct corresponding brain networks. Based on the existing literature (Binder et al., 2016; Fernandino et al., 2022; Lin et al., 2024; Tong et al., 2022), semantic representations are exemplars of such multidimensional representations. Other types of representations, such as visual representations, also exhibit multidimensionality and are compatible with the approach proposed in this study. However, in contrast to the whole-brain analysis method employed in the present work, visual functions may have more distinct and localized neural bases (Cox and Savoy, 2003; Ganis et al., 2004). Consequently, it may be more effective to first identify the brain regions implicated in the visual system and subsequently partition these regions into smaller subnetworks for more targeted analysis. Furthermore, the proposed method can be integrated with dynamic functional network approaches to investigate how information on the dynamic changes in target cognitive functions influences brain network reorganization. Specifically, task-induced brain activations can first be segmented into distinct stages using dynamic functional network approaches. Subsequently, the proposed method can be applied to the brain activations within each stage to obtain brain network partitions based on the multidimensional representations of the target cognitive functions.

Second, to evaluate the semantic similarity between computational models and the human brain (detailed in Section 2.4.3), we introduce a novel model-brain alignment method. First, we map the representations of computational models and brain activity into a shared semantic space, defined by 59 semantic dimensions. Subsequently, we compute the Pearson correlation between the representations of the computational models and brain activity within this common space. Previous encoding methods directly map hidden representations onto brain activation patterns and assess the mapping performance using the Pearson correlation coefficient (Schrimpf et al., 2021; A. Y. Wang et al., 2023). However, both brain and model representations encompass various types of information, including semantic content, grammatical structure, low-level features, and noise. As a result, a high correlation does not indicate that the two representations share similar semantic information. The proposed model-brain alignment method filters out noise and semantically irrelevant information, providing a more accurate reflection of the similarity between two sources of semantic information. Thus, researchers can use the proposed method to

further investigate the semantic similarity between human brain activity and computational models.

We applied the proposed method to construct a semantic brain network partition using fMRI data from a concept comprehension task. This network partition can be employed to interpret region-level semantic-related functions through fMRI, EEG, or MEG data. Furthermore, some studies focused on semantics concentrate only on specific brain regions, and incorporating the entire brain into the analysis can significantly increase both computational time and complexity. Our semantic network partition serves as a set of masks that preserve semantically relevant large-scale network units, thereby enhancing computational efficiency. Moreover, several studies on model-brain alignment have performed ROI-level analyses using resting-state network partitions (Sun et al., 2024; Y. Zhang et al., 2023b). In comparison to resting-state network partitions, our proposed partition not only enables researchers to exclude semantically irrelevant brain regions but also provides a deeper understanding of the semantic functions within each network. Researchers can select the appropriate semantic network based on their specific research objectives.

Then, we discussed the potential impacts of non-semantic task-related components on the proposed method. A considerable number of recent studies have employed multi-dimensional semantic decoding (or encoding) methods to investigate the neural correlates of semantic representations in the human brain, many of which utilize experimental paradigms without any contrast conditions (Caucheteux et al., 2023; Fernandino et al., 2022; Huth et al., 2016; Sun et al., 2020; Tong et al., 2022; S. Wang et al., 2024, 2022b; Y. Zhang et al., 2023a, 2023b; Zhang et al., 2024). For non-semantic components such as attention and memory retrieval, which are difficult to explicitly quantify, these studies face challenges in achieving strict control. However, researchers in the field of semantic representation generally hold an optimistic attitude towards the reliability of multidimensional semantic decoding analysis. According to our understanding, the methodological consideration is that multidimensional semantic decoding analysis should be much less sensitive to confounding by non-semantic components compared to traditional univariate analysis. Although the activation maps of words (i.e., the t-value images detailed in Section 2.1.1) represents activations triggered by both semantic and non-semantic components (such as attention and memory retrieval), multidimensional semantic decoding analysis only extracts information associated with the variations of semantic dimensions from them to reflect semantic representation. More importantly, multi-dimensional semantic decoding analysis integrates neural correlates across numerous semantic dimensions to reflect the semantic

representations, and the effects of a single non-semantic cognitive component are unlikely to systematically confound with multiple semantic dimensions. Therefore, current research generally employs multidimensional semantic decoding as a method to reflect semantic brain representation in no-contrast experimental paradigms.

The present study has several limitations. First, although the fMRI data we used is currently the largest open-source dataset available for concept comprehension, the number of participants is still smaller compared to resting-state datasets. To encompass a broad spectrum of concepts, the dataset developed by S. Wang et al. (2023) necessitated extended data collection periods for each participant, resulting in a smaller overall participant pool. Future research will seek to validate the applicability of our methodology using datasets with larger participant cohorts. Second, although our experiments demonstrate that the proposed method exhibits high stability across different models, the delineation of brain networks is still somewhat influenced by the model, especially when partitioning brain regions associated with weaker target cognitive functions. This influence can be reduced by first defining the brain areas related to the target cognitive functions and then subdividing these regions into smaller networks. Third, for the same target cognitive function, datasets from different tasks can influence the resulting brain network partitions. Although it is generally accepted that the semantic representation functions of brain regions (semantic memory) remain relatively stable across various tasks, the activation induced by semantic representations depends on specific stimuli and task demands. Therefore, the choice of stimuli and tasks influence both the type and extent of semantic activation. Fourth, we only delineate the cortical semantic partition. Previous studies have found that semantic processing primarily involves widely distributed cortical regions (Binder et al., 2009). However, recent studies have identified that some sub-cortical regions (e.g., the cerebellum, thalamus, and caudate nucleus) also contribute to verbal semantic comprehension (Cocquyt et al., 2019; Turker et al., 2023). Therefore, delineating sub-cortical semantic partitions will further clarify the semantic organization of sub-cortical structures.

5 Conclusions

We developed a method for constructing brain networks based on the homogeneity and heterogeneity of brain regions in specific cognitive functions, and demonstrated the reliability and validity of the method through semantic-network partitioning. Our findings indicate that our method can reliably partition semantic networks, and its results are distinctly different from those of traditional brain network parcellation. The results of brain semantic network parcellation

exhibit high stability and cross-semantic model consistency in brain regions representing semantic information. Different brain networks show systematic differences in their semantic-related functions. Our study provides a method for brain network parcellation in functional neuroimaging studies focusing on a specific type of cognitive function. In addition, our results also shed light on the organization of semantic functions in human brain.

Ethics statement

The cognitive data used in this study have been processed to ensure that they do not contain any information that can be directly linked to the participants' identities. The data were shared by the Institute of Automation, Chinese Academy of Sciences (CAS), and all experiments were approved by the Institutional Ethics Committee at the Institute of Psychology, Chinese Academy of Sciences, in accordance with ethical guidelines and regulations.

Declaration of Competing Interest

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

Credit authorship contribution statement

Yunhao Zhang: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Visualization, Project administration. **Shaonan Wang**: Conceptualization, Writing – original draft, Supervision. **Nan Lin**: Conceptualization, Methodology, Resources, Writing – original draft, review & editing, Supervision. **Lingzhong Fan**: Conceptualization, Methodology, Writing – review & editing. **Chengqing Zong**: Resources, Writing – review & editing, Funding acquisition

Data availability statement

The data for the paper is made publicly available on the following OpenNeuro page: https://openneuro.org/datasets/ds004301, and the following OSF page: https://osf.io/n5vke/. The code used to run the analyses and the proposed semantic network partition can be found on GitHub: https://github.com/Cupid777/Semantic-Network-Partition.

Acknowledgement

This research was supported by grants from the National Natural Science Foundation of China to S.W. (62036001) and S.W. (the STI2030-Major Project, grant number: 2021ZD0204105).

References

Adams, R., Bischof, L., 1994. Seeded region growing. IEEE Transactions on pattern analysis and machine intelligence 16, 641–647.

Anderson, A.J., Binder, J.R., Fernandino, L., Humphries, C.J., Conant, L.L., Aguilar, M., Wang, X., Doko, D., Raizada, R.D., 2017. Predicting neural activity patterns associated with sentences using a neurobiologically motivated model of semantic representation. Cerebral Cortex 27, 4379–4395.

Anderson, A.J., Zinszer, B.D., Raizada, R.D., 2016. Representational similarity encoding for fMRI: Pattern-based synthesis to predict brain activity using stimulus-model-similarities. NeuroImage 128, 44–53.

Arioli, M., Gianelli, C., Canessa, N., 2021. Neural representation of social concepts: a coordinate-based meta-analysis of fMRI studies. Brain Imaging and Behavior 15, 1912–1921. https://doi.org/10.1007/s11682-020-00384-6

Bao, H., Dong, L., Piao, S., Wei, F., 2022. BEiT: BERT Pre-Training of Image Transformers. https://doi.org/10.48550/arXiv.2106.08254

Bi, Y., 2021. Dual coding of knowledge in the human brain. Trends in Cognitive Sciences 25, 883–895.

Binder, J.R., Conant, L.L., Humphries, C.J., Fernandino, L., Simons, S.B., Aguilar, M., Desai, R.H., 2016. Toward a brain-based componential semantic representation. Cognitive Neuropsychology 33, 130–174. https://doi.org/10.1080/02643294.2016.1147426

Binder, J.R., Desai, R.H., 2011. The neurobiology of semantic memory. Trends in cognitive sciences 15, 527–536.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. Cerebral Cortex 19, 2767–2796. https://doi.org/10.1093/cercor/bhp055

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment 2008, P10008.

Bommasani, R., Davis, K., Cardie, C., 2020. Interpreting pretrained contextualized representations via reductions to static embeddings, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 4758–4781.

Bressler, S.L., Menon, V., 2010. Large-scale brain networks in cognition: emerging methods and principles. Trends in cognitive sciences 14, 277–290.

Bruurmijn, M.L., Pereboom, I.P., Vansteensel, M.J., Raemaekers, M.A., Ramsey, N.F., 2017. Preservation of hand movement representation in the sensorimotor areas of amputees. Brain 140, 3166–3178.

Cao, H., Plichta, M.M., Schäfer, A., Haddad, L., Grimm, O., Schneider, M., Esslinger, C., Kirsch, P., Meyer-Lindenberg, A., Tost, H., 2014. Test–retest reliability of fMRI-based graph theoretical properties during working memory, emotion processing, and resting state. Neuroimage 84, 888–900.

Carota, F., Kriegeskorte, N., Nili, H., Pulvermüller, F., 2017. Representational similarity mapping of distributional semantics in left inferior frontal, middle temporal, and motor cortex. Cerebral Cortex 27, 294–309.

Caucheteux, C., Gramfort, A., King, J.-R., 2023. Evidence of a predictive coding hierarchy in the human brain listening to speech. Nature human behaviour 7, 430–441.

Chen, C., Dupré la Tour, T., Gallant, J.L., Klein, D., Deniz, F., 2024. The cortical representation of language timescales is shared between reading and listening. Communications Biology 7, 284.

Chersoni, E., Santus, E., Huang, C.-R., Lenci, A., 2021. Decoding word embeddings with brain-based semantic features. Computational Linguistics 47, 663–698.

Cocquyt, E.-M., Coffé, C., van Mierlo, P., Duyck, W., Mariën, P., Szmalec, A., Santens, P., De Letter, M., 2019. The involvement of subcortical grey matter in verbal semantic comprehension: A systematic review and meta-analysis of fMRI and PET studies. Journal of Neurolinguistics 51, 278–296.

Cohen, A.D., Chen, Z., Parker Jones, O., Niu, C., Wang, Y., 2020. Regression-based machine-learning approaches to predict task activation using resting-state fMRI. Human Brain Mapping 41, 815–826. https://doi.org/10.1002/hbm.24841

Cole, M.W., Ito, T., Cocuzza, C., Sanchez-Romero, R., 2021. The functional relevance of task-state functional connectivity. Journal of Neuroscience 41, 2684–2702.

Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.-C., Abdi, H., Haxby, J.V., 2012. The representation of biological classes in the human brain. Journal of Neuroscience 32, 2608–2618.

Cox, D.D., Savoy, R.L., 2003. Functional magnetic resonance imaging (fMRI)"brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19, 261–270.

Cui, Y., Che, W., Liu, T., Qin, B., Wang, S., Hu, G., 2020. Revisiting Pre-Trained Models for Chinese Natural Language Processing, in: Findings of the Association for Computational Linguistics: EMNLP 2020. pp. 657–668. https://doi.org/10.18653/v1/2020.findings-emnlp.58

Cui, Y., Che, W., Liu, T., Qin, B., Yang, Z., 2021. Pre-training with whole word masking for chinese bert. IEEE/ACM Transactions on Audio, Speech, and Language Processing 29, 3504–3514.

de Zubicaray, G., Arciuli, J., McMahon, K., 2013. Putting an "end" to the motor cortex representations of action words. Journal of cognitive neuroscience 25, 1957–1974.

Deco, G., Tononi, G., Boly, M., Kringelbach, M.L., 2015. Rethinking segregation and integration: contributions of whole-brain modelling. Nature Reviews Neuroscience 16, 430–439.

Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53, 1–15.

Devereux, B.J., Clarke, A., Marouchos, A., Tyler, L.K., 2013. Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. Journal of Neuroscience 33, 18906–18916.

Devlin, J., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Dockès, J., Poldrack, R.A., Primet, R., Gözükan, H., Yarkoni, T., Suchanek, F., Thirion, B., Varoquaux, G., 2020. NeuroQuery, comprehensive meta-analysis of human brain mapping. elife 9, e53385.

Dosovitskiy, A., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Doucet, G.E., He, X., Sperling, M.R., Sharan, A., Tracy, J.I., 2017. From "rest" to language task: Task activation selects and prunes from broader resting-state network. Human Brain Mapping 38, 2540–2552. https://doi.org/10.1002/hbm.23539

Dreyer, F.R., Pulvermüller, F., 2018. Abstract semantics in the motor system?–An event-related fMRI study on passive reading of semantic word categories carrying abstract emotional and mental meaning. Cortex 100, 52–70.

Eickhoff, S.B., Yeo, B.T., Genon, S., 2018. Imaging-based parcellations of the human brain. Nature Reviews Neuroscience 19, 672–686.

Epstein, R.A., 2008. Parahippocampal and retrosplenial contributions to human spatial navigation. Trends in cognitive sciences 12, 388–396.

Epstein, R.A., Patai, E.Z., Julian, J.B., Spiers, H.J., 2017. The cognitive map in humans: spatial navigation and beyond. Nature neuroscience 20, 1504–1513.

Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., 2019. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature methods 16, 111–116.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., Fox, P.T., Eickhoff, S.B., Yu, C., Jiang, T., 2016. The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. Cereb. Cortex 26, 3508–3526. https://doi.org/10.1093/cercor/bhw157

Fernandino, L., Binder, J.R., Desai, R.H., Pendl, S.L., Humphries, C.J., Gross, W.L., Conant, L.L., Seidenberg, M.S., 2016. Concept representation reflects multimodal abstraction: A framework for embodied semantics. Cerebral cortex 26, 2018–2034.

Fernandino, L., Humphries, C.J., Seidenberg, M.S., Gross, W.L., Conant, L.L., Binder, J.R., 2015. Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. Neuropsychologia 76, 17–26.

Fernandino, L., Iacoboni, M., 2010. Are cortical motor maps based on body parts or coordinated actions? Implications for embodied semantics. Brain and language 112, 44–53.

Fernandino, L., Tong, J.-Q., Conant, L.L., Humphries, C.J., Binder, J.R., 2022. Decoding the information structure underlying the neural representation of concepts. Proc. Natl. Acad. Sci. U.S.A. 119, e2108091119. https://doi.org/10.1073/pnas.2108091119

Fischer, C.E., Churchill, N., Leggieri, M., Vuong, V., Tau, M., Fornazzari, L.R., Thaut, M.H., Schweizer, T.A., 2021. Long-known music exposure effects on brain imaging and cognition in early-stage cognitive decline: A pilot study. Journal of Alzheimer's Disease 84, 819–833.

Frisby, S.L., Halai, A.D., Cox, C.R., Ralph, M.A.L., Rogers, T.T., 2023. Decoding semantic representations in mind and brain. Trends in cognitive sciences 27, 258–281.

Fu, Z., Wang, Xiaosha, Wang, Xiaoying, Yang, H., Wang, J., Wei, T., Liao, X., Liu, Z., Chen, H., Bi, Y., 2023. Different computational relations in language are captured by distinct brain systems. Cerebral Cortex 33, 997–1013.

Ganis, G., Thompson, W.L., Kosslyn, S.M., 2004. Brain areas underlying visual mental imagery and visual perception: an fMRI study. Cognitive Brain Research 20, 226–241.

Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., 2016. A multi-modal parcellation of human cerebral cortex. Nature 536, 171–178.

Godwin, D., Barry, R.L., Marois, R., 2015. Breakdown of the brain's functional network modularity with awareness. Proc. Natl. Acad. Sci. U.S.A. 112, 3799–3804. https://doi.org/10.1073/pnas.1414466112

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S.A., Feder, A., Emanuel, D., Cohen, A., 2022. Shared computational principles for language processing in humans and deep language models. Nature neuroscience 25, 369–380.

Hein, G., Knight, R.T., 2008. Superior temporal sulcus—it's my area: or is it? Journal of cognitive neuroscience 20, 2125–2136.

Holland, R., Lambon Ralph, M.A., 2010. The anterior temporal lobe semantic hub is a part of the language neural network: selective disruption of irregular past tense verbs by rTMS. Cerebral Cortex 20, 2771–2775.

Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., Gallant, J.L., 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532, 453–458. https://doi.org/10.1038/nature17637

Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76, 1210–1224.

Jackson, R.L., Hoffman, P., Pobric, G., Ralph, M.A.L., 2016. The semantic network at work and rest: differential connectivity of anterior temporal lobe subregions. Journal of Neuroscience 36, 1490–1501.

Ji, J.L., Spronk, M., Kulkarni, K., Repovš, G., Anticevic, A., Cole, M.W., 2019. Mapping the human brain's cortical-subcortical functional network organization. Neuroimage 185, 35–57.

Jones, O.P., Voets, N.L., Adcock, J.E., Stacey, R., Jbabdi, S., 2017. Resting connectivity predicts task activation in pre-surgical populations. NeuroImage: Clinical 13, 378–385.

Kim, W., Son, B., Kim, I., 2021. Vilt: Vision-and-language transformer without convolution or region supervision, in: International Conference on Machine Learning. PMLR, pp. 5583–5594.

Kuhnke, P., Chapman, C.A., Cheung, V.K.M., Turker, S., Graessner, A., Martin, S., Williams, K.A., Hartwigsen, G., 2023. The role of the angular gyrus in semantic cognition: a synthesis of five functional neuroimaging studies. Brain Struct Funct 228, 273–291. https://doi.org/10.1007/s00429-022-02493-y

Kumar, A.A., 2021. Semantic memory: A review of methods, models, and current challenges. Psychon Bull Rev 28, 40–80. https://doi.org/10.3758/s13423-020-01792-x

Laird, A.R., Lancaster, J.L., Fox, P.T., 2005. BrainMap: The Social Evolution of a Human Brain Mapping Database. NI 3, 065–078. https://doi.org/10.1385/NI:3:1:065

Lenci, A., Lebani, G.E., Passaro, L.C., 2018. The Emotions of Abstract Words: A Distributional Semantic Analysis. Topics in Cognitive Science 10, 550–572. https://doi.org/10.1111/tops.12335

Li, C., Wang, S., Zhang, Y., Zhang, J., Zong, C., 2023. Interpreting and Exploiting Functional Specialization in Multi-Head Attention under Multi-task Learning. https://doi.org/10.48550/arXiv.2310.10318

Li, J., Tang, T., Zhao, W.X., Nie, J.-Y., Wen, J.-R., 2024. Pre-Trained Language Models for Text Generation: A Survey. ACM Comput. Surv. 56, 1–39. https://doi.org/10.1145/3649449

Li, X., Dvornek, N.C., Zhou, Y., Zhuang, J., Ventola, P., Duncan, J.S., 2019. Graph Neural Network for Interpreting Task-fMRI Biomarkers, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 485–493. https://doi.org/10.1007/978-3-030-32254-0_54

Lin, N., Wang, Xiaoying, Xu, Y., Wang, Xiaosha, Hua, H., Zhao, Y., Li, X., 2018a. Fine subdivisions of the semantic network supporting social and sensory–motor semantic processing. Cerebral Cortex 28, 2699–2710.

Lin, N., Xu, Y., Wang, X., Yang, H., Du, M., Hua, H., Li, X., 2019. Coin, telephone, and handcuffs: Neural correlates of social knowledge of inanimate objects. Neuropsychologia 133, 107187.

Lin, N., Xu, Y., Yang, H., Zhang, G., Zhang, M., Wang, S., Hua, H., Li, X., 2020. Dissociating the neural correlates of the sociality and plausibility effects in simple conceptual combination. Brain Struct Funct 225, 995–1008. https://doi.org/10.1007/s00429-020-02052-3

Lin, N., Yang, X., Li, J., Wang, S., Hua, H., Ma, Y., Li, X., 2018b. Neural correlates of three cognitive processes involved in theory of mind and discourse comprehension. Cogn Affect Behav Neurosci 18, 273–283. https://doi.org/10.3758/s13415-018-0568-6

Lin, N., Zhang, X., Wang, X., Wang, S., 2024. The organization of the semantic network as reflected by the neural correlates of six semantic dimensions. Brain and Language 250, 105388.

Lu, Y., Jiang, T., Zang, Y., 2003. Region growing method for the analysis of functional MRI data. NeuroImage 20, 455–465.

Ludersdorfer, P., Price, C.J., Duncan, K.J.K., DeDuck, K., Neufeld, N.H., Seghier, M.L., 2019. Dissociating the functions of superior and inferior parts of the left ventral occipito-temporal cortex during visual word and object processing. NeuroImage 199, 325–335.

Luo, Y., Xu, M., Xiong, D., 2022. CogTaskonomy: Cognitively Inspired Task Taxonomy Is Beneficial to Transfer Learning in NLP, in: Muresan, S., Nakov, P., Villavicencio, A. (Eds.), Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Dublin, Ireland, pp. 904–920. https://doi.org/10.18653/v1/2022.acl-long.64

Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., Jefferies, E., Smallwood, J., 2016. Situating

the default-mode network along a principal gradient of macroscale cortical organization. Proc. Natl. Acad. Sci. U.S.A. 113, 12574–12579. https://doi.org/10.1073/pnas.1608282113

Martin, A., 2007. The Representation of Object Concepts in the Brain. Annu. Rev. Psychol. 58, 25–45. https://doi.org/10.1146/annurev.psych.57.102904.190143

Mattheiss, S.R., Levinson, H., Graves, W.W., 2018. Duality of function: activation for meaningless nonwords and semantic codes in the same brain areas. Cerebral Cortex 28, 2516–2524.

Moraschi, M., Mascali, D., Tommasin, S., Gili, T., Hassan, I.E., Fratini, M., DiNuzzo, M., Wise, R.G., Mangia, S., Macaluso, E., 2020. Brain network modularity during a sustained working-memory task. Frontiers in Physiology 11, 422.

Ngo, G.H., Eickhoff, S.B., Nguyen, M., Sevinc, G., Fox, P.T., Spreng, R.N., Yeo, B.T., 2019. Beyond consensus: embracing heterogeneity in curated neuroimaging meta-analysis. Neuroimage 200, 142–158.

Ngo, G.H., Khosla, M., Jamison, K., Kuceyeski, A., Sabuncu, M.R., 2022a. Predicting individual task contrasts from resting-state functional connectivity using a surface-based convolutional network. NeuroImage 248, 118849.

Ngo, G.H., Nguyen, M., Chen, N.F., Sabuncu, M.R., 2022b. A transformer-Based neural language model that synthesizes brain activation maps from free-form text queries. Medical image analysis 81, 102540.

Niu, C., Wang, Y., Cohen, A.D., Liu, X., Li, H., Lin, P., Chen, Z., Min, Z., Li, W., Ling, X., Wen, X., Wang, M., Thompson, H.P., Zhang, M., 2021. Machine learning may predict individual hand motor activation from resting-state fMRI in patients with brain tumors in perirolandic cortex. Eur Radiol 31, 5253–5262. https://doi.org/10.1007/s00330-021-07825-w

Oota, S., Gupta, M., Toneva, M., 2024. Joint processing of linguistic properties in brains and language models. Advances in Neural Information Processing Systems 36.

Paivio, A., 1990. Mental representations: A dual coding approach. Oxford university press.

Patel, E., Kushwaha, D.S., 2020. Clustering cloud workloads: K-means vs gaussian mixture model. Procedia computer science 171, 158–167.

Patel, G.H., Arkin, S.C., Ruiz-Betancourt, D.R., Plaza, F.I., Mirza, S.A., Vieira, D.J., Strauss, N.E., Klim, C.C., Sanchez-Peña, J.P., Bartel, L.P., 2021. Failure to engage the temporoparietal junction/posterior superior temporal sulcus predicts impaired naturalistic social cognition in schizophrenia. Brain 144, 1898–1910.

Pessoa, L., 2014. Understanding brain networks and brain organization. Physics of life reviews 11, 400–435.

Petersen, S.E., Sporns, O., 2015. Brain networks and cognitive architectures. Neuron 88, 207–219.

Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., 2011. Functional network organization of the human brain. Neuron 72, 665–678.

Power, J.D., Fair, D.A., Schlaggar, B.L., Petersen, S.E., 2010. The development of human functional brain networks. Neuron 67, 735–748.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., 2019. Language models are unsupervised multitask learners. OpenAI blog 1, 9.

Recchia, G., Louwerse, M.M., 2015. Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. Quarterly Journal of Experimental Psychology 68, 1584–1598. https://doi.org/10.1080/17470218.2014.941296

Rolinski, R., You, X., Gonzalez-Castillo, J., Norato, G., Reynolds, R.C., Inati, S.K., Theodore, W.H., 2020. Language lateralization from task-based and resting state functional MRI in patients with epilepsy. Human Brain Mapping 41, 3133–3146. https://doi.org/10.1002/hbm.25003

Rubinov, M., Sporns, O., 2011. Weight-conserving characterization of complex functional brain networks. Neuroimage 56, 2068–2079.

Saarimäki, H., 2021. Naturalistic stimuli in affective neuroimaging: A review. Frontiers in human neuroscience 15, 675068.

Salehi, M., Greene, A.S., Karbasi, A., Shen, X., Scheinost, D., Constable, R.T., 2020. There is no single functional atlas even for a single individual: Functional parcel definitions change with task. NeuroImage 208, 116366.

Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T., 2018. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. Cerebral cortex 28, 3095–3114.

Schneider, H.R., Wawrzyniak, M., Stockert, A., Klingbeil, J., Saur, D., 2022. fMRI informed voxel-based lesion analysis to identify lesions associated with right-hemispheric activation in aphasia recovery. NeuroImage: Clinical 36, 103169.

Schrimpf, M., Blank, I.A., Tuckute, G., Kauf, C., Hosseini, E.A., Kanwisher, N., Tenenbaum, J.B., Fedorenko, E., 2021. The neural architecture of language: Integrative modeling converges on predictive processing. Proc. Natl. Acad. Sci. U.S.A. 118, e2105646118. https://doi.org/10.1073/pnas.2105646118

Su, J., Ahmed, M., Lu, Y., Pan, S., Bo, W., Liu, Y., 2024. Roformer: Enhanced transformer with rotary position embedding. Neurocomputing 568, 127063.

Sun, J., Li, M., Chen, Z., Zhang, Y., Wang, S., Moens, M.-F., 2024. Contrast, attend and diffuse to decode high-resolution images from brain activities. Advances in Neural Information Processing Systems 36.

Sun, J., Wang, S., Zhang, J., Zong, C., 2020. Neural encoding and decoding with distributed sentence representations. IEEE Transactions on Neural Networks and Learning Systems 32, 589–603.

Sun, Y., Wang, S., Feng, S., Ding, S., Pang, C., Shang, J., Liu, J., Chen, X., Zhao, Y., Lu, Y., Liu, Weixin, Wu, Z., Gong, W., Liang, J., Shang, Z., Sun, P., Liu, Wei, Ouyang, X., Yu, D., Tian, H., Wu, H., Wang, H., 2021. ERNIE 3.0: Large-scale Knowledge Enhanced Pre-training for Language Understanding and Generation. https://doi.org/10.48550/arXiv.2107.02137

Tavor, I., Jones, O.P., Mars, R.B., Smith, S.M., Behrens, T.E., Jbabdi, S., 2016. Task-free MRI predicts individual differences in brain activity during task performance. Science 352, 216–220. https://doi.org/10.1126/science.aad8127

Thomas Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., Fischl, B., Liu, H., Buckner, R.L., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology 106, 1125–1165. https://doi.org/10.1152/jn.00338.2011

Toneva, M., Mitchell, T.M., Wehbe, L., 2022. Combining computational controls with natural text reveals aspects of meaning composition. Nature computational science 2, 745–757.

Toneva, M., Wehbe, L., 2019. Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain). Advances in neural information processing systems 32.

Tong, J., Binder, J.R., Humphries, C., Mazurchuk, S., Conant, L.L., Fernandino, L., 2022. A distributed network for multimodal experiential representation of concepts. Journal of Neuroscience 42, 7121–7130.

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jégou, H., 2021. Training data-efficient image transformers & distillation through attention, in: International Conference on Machine Learning. PMLR, pp. 10347–10357.

Turker, S., Kuhnke, P., Eickhoff, S.B., Caspers, S., Hartwigsen, G., 2023. Cortical, subcortical, and cerebellar contributions to language processing: A meta-analytic review of 403 neuroimaging experiments. Psychological Bulletin.

Utsumi, A., 2020. Exploring What Is Encoded in Distributional Word Vectors: A Neurobiologically Motivated Analysis. Cognitive Science 44, e12844. https://doi.org/10.1111/cogs.12844

Vulić, I., Ponti, E.M., Litschko, R., Glavaš, G., Korhonen, A., 2020. Probing pretrained language models for lexical semantics, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 7222–7240.

Wang, A.Y., Kay, K., Naselaris, T., Tarr, M.J., Wehbe, L., 2023. Better models of human high-level visual cortex emerge from natural language supervision with a large and diverse dataset. Nature Machine Intelligence 5, 1415–1426.

Wang, D., Buckner, R.L., Fox, M.D., Holt, D.J., Holmes, A.J., Stoecklein, S., Langs, G., Pan, R., Qian, T., Li, K., 2015. Parcellating cortical functional networks in individuals. Nature neuroscience 18, 1853–1860.

Wang, P., Yang, A., Men, R., Lin, J., Bai, S., Li, Z., Ma, J., Zhou, C., Zhou, J., Yang, H., 2022. Ofa: Unifying architectures, tasks, and modalities through a simple sequence-to-sequence learning framework, in: International Conference on Machine Learning. PMLR, pp. 23318–23340.

Wang, S., Sun, J., Zhang, Y., Lin, N., Moens, M.-F., Zong, C., 2024. Computational Models to Study Language Processing in the Human Brain: A Survey. https://doi.org/10.48550/arXiv.2403.13368

Wang, S., Zhang, X., Zhang, J., Zong, C., 2022a. A synchronized multimodal neuroimaging dataset for studying brain language processing. Scientific Data 9, 590.

Wang, S., Zhang, Y., Shi, W., Zhang, G., Zhang, J., Lin, N., Zong, C., 2023. A large dataset of semantic ratings and its computational extension. Scientific Data 10, 106.

Wang, S., Zhang, Y., Zhang, X., Sun, J., Lin, N., Zhang, J., Zong, C., 2022b. An fmri dataset for concept representation with semantic feature annotations. Scientific Data 9, 721.

Wang, Y., Gong, X., Meng, L., Wu, X., Meng, H., 2024. Large Language Model-based FMRI Encoding of Language Functions for Subjects with Neurocognitive Disorder, in: Proc. Interspeech 2024. pp. 1485–1489.

Xu, Y., Lin, Q., Han, Z., He, Y., Bi, Y., 2016. Intrinsic functional network architecture of human semantic processing: Modules and hubs. Neuroimage 132, 542–555.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nature methods 8, 665–670.

Yee, E., Thompson-Schill, S.L., 2016. Putting concepts into context. Psychon Bull Rev 23, 1015–1027. https://doi.org/10.3758/s13423-015-0948-7

Yeo, B.T., Krienen, F.M., Eickhoff, S.B., Yaakub, S.N., Fox, P.T., Buckner, R.L., Asplund, C.L., Chee, M.W., 2015. Functional specialization and flexibility in human association cortex. Cerebral cortex 25, 3654–3672.

Zhang, G., Hung, J., Lin, N., 2023. Coexistence of the social semantic effect and non-semantic effect in the default mode network. Brain Struct Funct 228, 321–339. https://doi.org/10.1007/s00429-022-02476-z

Zhang, J., Gan, R., Wang, J., Zhang, Y., Zhang, L., Yang, P., Gao, X., Wu, Z., Dong, X., He, J., Zhuo, J., Yang, Q., Huang, Y., Li, X., Wu, Y., Lu, J., Zhu, X., Chen, W., Han, T., Pan, K., Wang, R., Wang, H., Wu, X., Zeng, Z., Chen, C., 2023. Fengshenbang 1.0: Being the Foundation of Chinese Cognitive Intelligence. https://doi.org/10.48550/arXiv.2209.02970

Zhang, Y., Li, C., Zhang, X., Dong, X., Wang, S., 2023a. A comprehensive neural and behavioral task taxonomy method for transfer learning in nlp, in: Findings of the Association for Computational Linguistics: IJCNLP-AACL 2023 (Findings). pp. 233–241.

Zhang, Y., Wang, S., Dong, X., Yu, J., Zong, C., 2023b. Navigating Brain Language Representations: A Comparative Analysis of Neural Language Models and Psychologically Plausible Models, in: Proceedings of the Annual Meeting of the Cognitive Science Society.

Zhang, Y., Zhang, X., Li, C., Wang, S., Zong, C., 2024. MulCogBench: A Multi-modal Cognitive Benchmark Dataset for Evaluating Chinese and English Computational Language Models. https://doi.org/10.48550/arXiv.2403.01116

Zhang, Z., Zhang, H., Chen, K., Guo, Y., Hua, J., Wang, Y., Zhou, M., 2021. Mengzi: Towards Lightweight yet Ingenious Pre-trained Models for Chinese. https://doi.org/10.48550/arXiv.2110.06696

Zhao, Y., Song, L., Ding, J., Lin, N., Wang, Q., Du, X., Sun, R., Han, Z., 2017. Left anterior temporal lobe and bilateral anterior cingulate cortex are semantic hub regions: Evidence from behavior-nodal degree mapping in brain-damaged patients. Journal of Neuroscience 37, 141–151.

Zhao, Z., Chen, H., Zhang, J., Zhao, X., Liu, T., Lu, W., Chen, X., Deng, H., Ju, Q., Du, X., 2019. UER: An Open-Source Toolkit for Pre-training Models. https://doi.org/10.48550/arXiv.1909.05658

Data and Code Availability Statement

The data for the paper is made publicly available on the following OpenNeuro page: https://openneuro.org/datasets/ds004301, and the following OSF page: https://osf.io/n5vke/. The code used to run the analyses and the proposed semantic network partition can be found on GitHub: https://github.com/Cupid777/Semantic-Network-Partition.

Declaration of Interest Statement

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