Topological and Modular Analysis of Semantic Networks in Alzheimer's Disease: Community Structure, Modularity, and Network Reorganization

Reproduction and Extension of Zemla & Austerweil (2019)

Authors: Gemma Esteve Valls, David Hidalgo Fàbregas, Zaid Al Hakioui El Mettioui

Group: Group A3N

Abstract

Alzheimer's Disease (AD) is a neurodegenerative condition affecting millions of older adults worldwide, yet early diagnosis and a full understanding of its cognitive impact remain challenging. This study reproduces and extends the work of Zemla and Austerweil (2019) by investigating the topological and modular properties of semantic networks in individuals with AD and cognitively normal controls (NC). Using data from semantic fluency tasks, we applied graph theory metrics and community detection algorithms (Louvain, Infomap, DCSBM) to examine network organization, modularity, and reorganization. Results indicate that AD networks are smaller, less modular, and denser due to reduced lexical output. Transition networks at the word and subcategory levels revealed impaired semantic navigation in AD. A Subject Similarity Network (SSN) approach enabled unsupervised grouping with over 80% accuracy, while machine learning classifiers achieved up to 87.3% accuracy. These findings highlight the diagnostic potential of semantic network analysis for early AD detection.

1. Introduction

1.1. Alzheimer’s disease and semantic memory deterioration

Alzheimer’s disease (AD) is a neurodegenerative condition characterized by a gradual decline in cognitive functions, affecting 46 millions of individuals worldwide [2]. At the biological level, the disease is associated with progressive neuronal damage and inconsistent communication between brain regions, which ultimately compromises higher cognitive processes. Deterioration of semantic memory is among the earliest and most prominent cognitive impairments observed in AD [3].

Semantic memory is responsible for structured knowledge about concepts, categories, and the relationships between them, forming the basis for meaningful language use. As this system degrades, patients experience increasing difficulty accessing and retrieving conceptual information, often manifesting as impaired word retrieval and reduced semantic coherence. These deficits are commonly observed even at early stages of the disease, making semantic memory a particularly relevant domain for both clinical assessment and research.

In clinical practice, semantic fluency tasks, such as asking individuals to name as many animals as possible within a limited time, are widely used to evaluate semantic memory integrity. While these tasks are easy to administer and sensitive to cognitive decline, their interpretation typically relies on basic performance measures, including the number of words produced or the occurrence of repetitions. Although informative, such measures offer limited insight into how semantic knowledge is internally organized, which may fail to capture more subtle changes in the structure of semantic memory caused by the disease.

1.2. Network theory metrics for cognitive systems

Network theory provides a mathematical framework for representing complex systems composed of interacting elements. In this framework, a system is modeled as a graph $G = (N, E)$, where N defines the set of nodes and E the set of edges describing interactions between them. When applied to cognitive systems, nodes can represent concepts or words, while edges encode semantic associations between them [4].

The structure of a network can be described through its adjacency matrix A, where each entry A_{ij} indicates the presence or absence of a connection between nodes i and j. In the present work, networks are treated as undirected and unweighted, resulting in a symmetric binary adjacency matrix.

Several topological metrics are commonly used to characterize network organization. The degree k_i of a node measures the number of connections it has and is defined as

$$k_i = \sum_j A_{ij}$$

Beyond local connectivity, network efficiency is captured through path-based measures. The shortest path length d_{ij} between two nodes corresponds to the minimum number of edges required to connect them, and the average shortest path length reflects how efficiently information can be exchanged across the network. Finally, the clustering coefficient measures the tendency of nodes to form locally interconnected neighborhoods, capturing the presence of local redundancy and modular structure.

In addition to these measures, several global descriptors are commonly used to characterize the overall organization of a network. Network density quantifies how many edges are present relative to the maximum possible number of edges and is defined as

$$\rho = \frac{2|E|}{N(N-1)},$$

where $|E|$ is the number of edges and N the number of nodes in the network. Density provides a normalized measure of connectivity that facilitates comparisons between networks of different sizes.

The local clustering coefficient of a node i is defined as

$$C_i = \frac{2T_i}{k_i(k_i-1)},$$

where k_i is the degree of node i and T_i denotes the number of triangles formed by neighbors of i. The average clustering coefficient is obtained by averaging C_i over all nodes and reflects the prevalence of locally clustered structures across the network.

Global integration is further characterized by the average shortest path length, defined as

$$\hat{l} = \frac{\sum_{i \neq j} d_{ij}}{N(N-1)},$$

where d_{ij} denotes the shortest path length between nodes i and j. Closely related to this measure, the network diameter is defined as the maximum shortest path length between any pair of nodes, providing an estimate of the network's overall spatial extension.

1.3. Small-world organization and efficiency in semantic networks

Many real-world networks exhibit a characteristic organization known as small-world structure, defined by the coexistence of high local clustering and short average path lengths. This configuration allows efficient global communication while preserving local specialization [5].

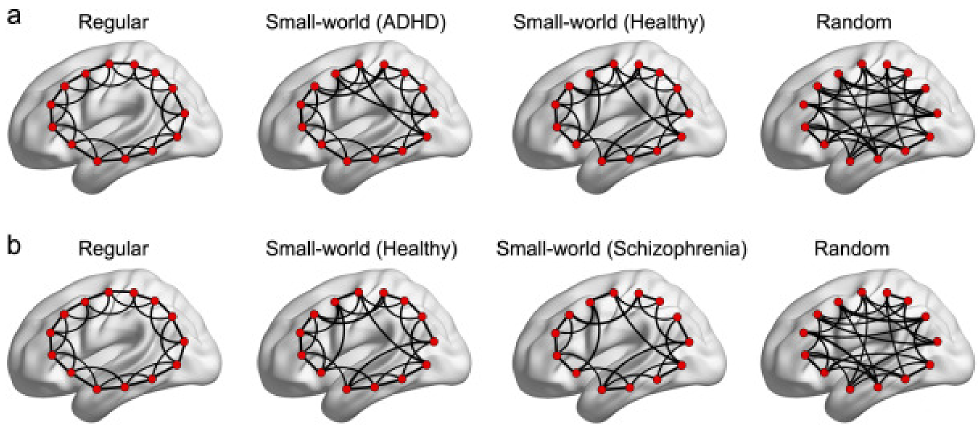


Figure 1. *Small-world human brain networks [5].*

Small-world networks occupy an intermediate regime between regular lattices, which are highly clustered but inefficient, and random networks, which are efficient but lack local structure. Formally, a network is considered small-world if its clustering coefficient is significantly higher than that of a random network of comparable size, while its average shortest path length remains similar.

This property has been observed in a wide range of systems, including biological, social, and linguistic networks. In particular, semantic networks derived from human language have consistently been shown to display small-world organization, suggesting that semantic memory is structured to balance efficient retrieval with meaningful conceptual grouping [6].

Small-worldness can be quantified using the coefficient σ, defined as:

$$\sigma = \frac{C/C_{rand}}{\hat{l}/\hat{l}_{rand}},$$

where C and $\hat{l}$ denote the clustering coefficient and average shortest path length of the original network, and C_{rand} and $\hat{l}_{rand}$ correspond to the same metrics averaged over an ensemble of randomized networks.

1.4. Semantic fluency tasks and semantic network reconstruction

Semantic fluency tasks provide an indirect window into the organization of semantic memory. During these tasks, individuals generate sequences of words belonging to a given semantic category, such as animals, under limited time. The order in which words are produced is not arbitrary, but reflects underlying semantic associations between concepts.

From a network perspective, words can be represented as nodes, while semantic associations define edges between them. This representation allows semantic memory to be studied as a structured system, where patterns of connectivity encode conceptual organization (see Figure 2).

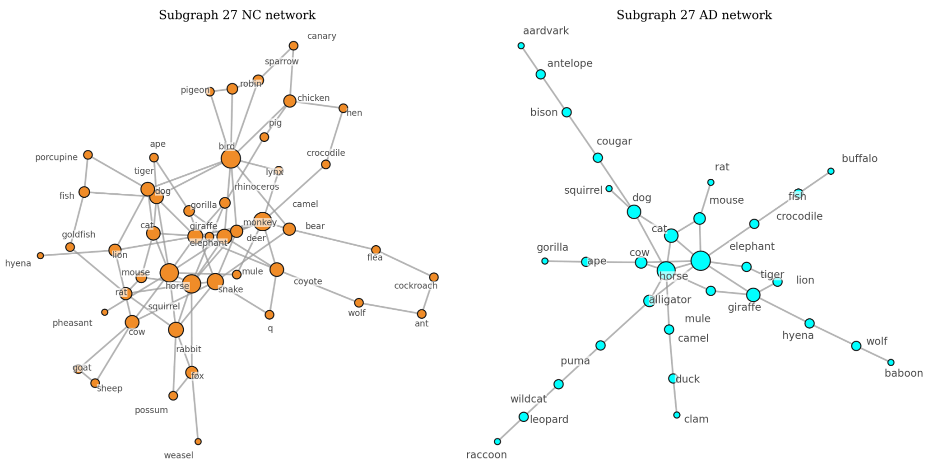


Figure 2. *Example: subgraph 27 for NC and AD networks.*

Building on this idea, semantic fluency data can be used not only to count how many words are produced, but also to infer how concepts are connected in memory. In the original work by *Zemla et al. (2019)*, the sequences of words generated by each participant were used to estimate individual semantic networks by assuming that people mentally move from one concept to another following existing semantic links. In simple terms, words that tend to be mentioned close together across fluency lists are interpreted as being more strongly related. This approach makes it possible to construct a network representation of semantic memory for each individual, capturing differences in how knowledge is organized rather than averaging behavior across groups.

1.5. Community structure in semantic memory

In network science, a community is generally defined as a group of nodes that are more densely connected to each other than to the rest of the network. This structural property reflects the tendency of elements within the same community to interact more frequently or more strongly, forming cohesive substructures embedded within the global network. In semantic networks, communities often correspond to conceptually related groups of words, such as semantic categories or subcategories, and their identification provides insight into how knowledge is organized and accessed.

From a methodological perspective, community structure can be quantified through the concept of modularity Q, which measures the extent to which a network can be partitioned into modules with dense internal connectivity and sparse external connections.

High modularity values indicate a clear separation between communities, whereas lower values suggest blurred boundaries between semantic categories. Approaches based in modularity rely on comparing the observed network structure to a null model, typically preserving node degrees, and aim to identify partitions that maximize within-community connectivity beyond what would be expected by chance. In the context of semantic networks, we think that modularity provides a global measure of how well-defined semantic categories are represented in the network.

Different community detection approaches rely on different conceptual principles. Louvain, for example, is a modularity-optimization method that seeks partitions with high within-community connectivity compared to a null model, using a greedy and hierarchical procedure that alternates local node reassignment with community aggregation, making it highly scalable for large networks [7]. Infomap, in contrast, frames community detection as an information-theoretic compression problem, identifying modules by minimizing the description length of a random walk on the network [8]. As a result, Infomap is particularly sensitive to flow-based structures and often detects smaller, more detailed communities, although it can be more sensitive to network density and local connectivity patterns.

In contrast, generative models such as the Degree-Corrected Stochastic Block Model (DCSBM) explicitly model community structure by assuming that edges are generated according to group-level connection probabilities while accounting for degree heterogeneity [9]. While Louvain and Infomap focus on optimizing structural (modularity) or informational (compression) criteria, DCSBM provides a probabilistic framework that helps disentangle community structure from degree effects.

1.6. Subjects Similarity Network

Once metrics from the semantic networks are obtained, we propose a data-driven classification approach. From the significant metrics of the analyzed networks, we tried to construct a Subjects Similarity Network (SNN), where nodes represent the subject and the edges represent the similarity between their topological profiles. To compute the similarity between subjects we computed the Euclidean distance, a metric that measures the straight-line distance between two points in the multidimensional feature space, sensitive to both the magnitude and the pattern of the topological metrics [10]. Given two feature vectors A and B for two subjects, the distance is defined as:

$$d(A,B) = \sqrt{\sum_{i=1}^{N} (A_i - B_i)^2},$$

where A_i corresponds to feature i from subject A, B_i corresponds to feature i from subject B, and N corresponds to the number of features to be considered in the feature vectors.

By constructing this network, we can apply community detection algorithms to identify clusters of subjects with similar cognitive network organizations. If the topological metrics derived from the fluency task are truly diagnostic, the SSN should naturally partition into communities that align with the clinical diagnoses (AD and NC), offering a validation of the disease groups without using labels during the construction process.

1.7. Proposed approach: Network Reorganization

One of the objectives of this project was to propose an alternative and simpler way to construct networks from animal semantic fluency lists that can effectively distinguish AD patients from NC.

Alzheimer's disease does not primarily involve an early degradation of semantic knowledge; rather, it disrupts controlled retrieval and strategic search processes, altering how individuals navigate semantic space (e.g., reduced switching in semantic fluency). Clustering and switching analyses show that AD patients tend to remain within local semantic patches and fail to switch efficiently between subcategories, reflecting deficits in strategic search rather than a pure loss of semantic representations [11]. Consequently, AD patients are expected to exhibit repeated short loops in fluency lists, become trapped within local subcategories, and fail to make long-range transitions across animal subcategories.

Traditional semantic network models (e.g., Zemla & Austerweil) define nodes as the words produced and edges based on semantic similarity norms, while the fluency list is used only to determine which nodes are included in the network. As a result, the temporal order in which words are retrieved is not incorporated into the network structure. In contrast, transition graph approaches derived from retrieval sequences explicitly integrate the temporal dynamics of semantic search, capturing process-level features that may provide additional discrimination between AD patients and controls.

In line with this perspective, our proposal constructs networks directly from the temporal dynamics of word retrieval, focusing on the navigation of semantic memory rather than its static structure. In the first approach, individual words are treated as nodes and directed edges represent successive retrievals, producing a network that explicitly captures word-order dynamics. In the second approach, animal names are classified into semantic subcategories, and networks are built from transitions between subcategories, emphasizing higher-level semantic navigation patterns.

2. Methodology

2.1. Dataset and Study Design

The dataset analyzed comes from the *National Alzheimer's Coordinating Center* (NACC) repository. The study includes a total of 158 subjects, classified into two experimental groups:

- AD Group (Alzheimer's Disease): Consisting of patients diagnosed with Alzheimer's disease.
- NC Group (Normal Control): Consisting of healthy control subjects.

The study design is cross-sectional and comparative, focused on identifying significant differences in the topology of the functional brain network between the two groups.

This work is based on the replication of the original study by *Zemla et al. (2019)*. To this end, this study starts with the pre-processed files provided by the authors of the reference study, which included both the adjacency matrices of the empirical functional networks and a set of 50 pre-calculated shuffled networks for each subject.

2.2. Network Representation and Preprocessing

The networks analyzed correspond to semantic graphs constructed from the results of the Animal Fluency Test. The network is modeled as a graph where the topology reflects the structure of the subject's semantic memory retrieval:

- Nodes (N): Each node represents a unique word (animal name) said by the subject during the test.
- Edges (E): A link is established between two nodes if the corresponding words were said consecutively (stimulus-response) in the temporal sequence of the test.

For the topological analysis, the provided adjacency matrices were used as a starting point. Although verbal production is intrinsically directional (time series), the networks were modeled as undirected and unweighted (binary) graphs to capture the underlying association structure independently of the retrieval order.

The preprocessing of the matrices included three main steps. Binarization was computed to ensure that all existing connections had a unit weight ($A_{ij} = 1$). Auto-loop removal was applied to delete immediate perseverations and avoid repeating the same word consecutively by forcing $A_{ij} = 0$. Finally, symmetrization was ensured to treat the relationship between nodes as bidirectional ($A_{ij} = A_{ji}$), assuming that the semantic association between two concepts is reciprocal in the network structure.

2.3. Global Network Metrics

Topological characterization was performed using the *NetworkX v3.6.1* library in *Python v3.13*. For each subject, the following global metrics were calculated:

- Basic Connectivity Metrics: The number of nodes (N), number of edges (E), average degree (k), and density (ρ) of the network were computed.
- Segregation: This was quantified using the Global Clustering Coefficient (C).
- Integration: This was evaluated using the Average Shortest Path Length ($\hat{l}$) and the diameter.

Since metrics such as Average Shortest Path Length and Diameter diverge in disconnected graphs, a connectivity validation step was implemented. If a network shows fragmentation, the Largest Connected Component (LCC) is extracted and the distance metrics are calculated exclusively on this main subgraph.

All metrics from each network were then compared among real vs shuffled network types to ensure that the results were due to network topology and not random chane. Finally, metric comparisons were made between NC and AD groups to draw conclusions.

2.4. Small-Worldness Analysis

To evaluate the Small-World (σ), the 50 randomly shuffled networks provided for each subject were used, rather than generating new simulations. These null networks were obtained by randomly permuting the order of items within each original list and then re-estimating the network using the same inference procedure as for the empirical network. This approach preserves the number of nodes and the overall item frequencies, while removing sequential dependencies between items. Small-World measures were then computed by comparing the empirical network to the distribution obtained from these shuffled networks. The comparative analysis was performed following the next steps:

- Reference Calculation: The Clustering Coefficient (C_{rand}) and Average Shortest Path Length ($\hat{l}_{rand}$) were computed for each of the 50 shuffled networks associated with a subject, and then both average values were obtained.
- Normalization: The metrics of the real network were normalized with respect to these averages.

$$\sigma = \frac{\gamma}{\lambda} = \frac{C_{real}/C_{rand}}{\hat{l}_{real}/\hat{l}_{rand}}$$

Finally, metric comparisons were made between NC and AD groups to draw conclusions.

2.5. Community Analysis

Firstly, to assess whether the observed semantic networks reflect genuine cognitive organization or are artifacts of the network size and density, we compared the following community metrics of the empirical networks against a set of null models applying Louvain, DCSBM and Infomap algorithms:

- Number of communities
- Average community size
- Modularity (Q)

For each subject, we used the 50 randomized networks that were generated by the authors of the paper, shuffling the edges while preserving the original degree distribution.

Then, the metric results of the real semantic networks were used to compare Normal Control (NC) versus Alzheimer's Disease (AD) groups to determine which parameters characterize the network depending on the type of subject.

Modularity was also used as a metric to evaluate the Louvain algorithm performance, as well as the Description Length (DL) was computed and used to evaluate the DCSBM algorithm, and the Code Length (CL) to evaluate the Infomap algorithm among both groups.

Finally, all metrics from each network were then compared among real vs shuffled network types to ensure that the results were due to network topology and not random chane. Finally, metric comparisons were made between NC and AD groups to draw conclusions.

2.6. Subjects Similarity Network

To investigate whether subjects naturally cluster by diagnosis based on their semantic network topology, we constructed a Subjects Similarity Network (SSN).

For each subject, we constructed a feature vector containing the normalized topological metrics calculated previously. The chosen features to be taken into account were the ones that were statistically significant when comparing the global and community metrics of AD among NC groups.

To assess the statistical significance of the topological differences observed, we employed a parametric t-test. First, to validate whether the empirical semantic networks exhibited non-random structural properties, we compared the topological metrics of each subject's real network against the average metrics of their corresponding 50 shuffled null model. For this comparison, paired sample t-tests were conducted, as the null models were generated directly from each subject's original data.

Second, to evaluate the impact of the pathology on network organization, we compared the metric distributions between the NC and AD using independent sample t-tests. Statistical significance was established at $p < 0.05$ for all comparisons.

Then, we computed a similarity matrix between all pairs of subjects using the Euclidean distance between their significant feature vectors. Based on these distances, we constructed the SSN as a k-nearest neighbors graph. In this graph, each node represents a subject, and an edge is created connecting a subject to its $k = 5$ closest peers in the topological feature space. This procedure filters out noisy connections.

Finally, we applied community detection algorithms (Louvain, Infomap and DCSBM) to this graph to identify clusters of subjects and evaluate their correspondence with the clinical diagnoses. As these algorithms do not permit us to condition the partition detection to 2 communities, we also performed Spectral Clustering and Agglomerative Clustering algorithms.

2.7. Network Reorganization and Exploratory Analysis

To redefine the networks based on the animal fluency lists we proceed the following pipeline.

First the fluency lists were loaded from the available data of the *Zemla et al. (2019)* repository. Animal fluency lists were obtained with participants classified as either cognitively normal or diagnosed with Alzheimer's disease. Each list consists of the words produced by a participant during a single fluency trial.

```
AD fluency list example: {0: 'giraffe', 1: 'kangaroo', 2: 'tiger', 3:
'lion', 4: 'alligator', 5: 'dog', 6: 'cat', 7: 'rabbit', 8: 'reptile',
9: 'zebra', 10: 'camel', 11: 'dromedary', 12: 'mouse'}.
```

Figure 3. *AD fluency list example.*

2.7.1. Construction of Transition Networks from Semantic Fluency Lists

Two types of transition (retrieval-dynamics) networks were generated from the animal semantic fluency lists in order to capture retrieval dynamics at different semantic levels: a word-level network and a subcategory-level network.

Word-level transition network

In the word-level approach, networks were constructed directly from the sequence of words produced by each participant. Each unique word corresponds to a node in the network. Directed edges represent the temporal order of retrieval, such that an edge from word A to word B indicates that B was produced immediately after A in the fluency list. Repeated word productions were retained, allowing the network to capture short recurrent loops and self-transitions that reflect retrieval difficulties frequently observed in Alzheimer's disease.

This representation preserves fine-grained word-order dynamics and local search behavior within semantic space.

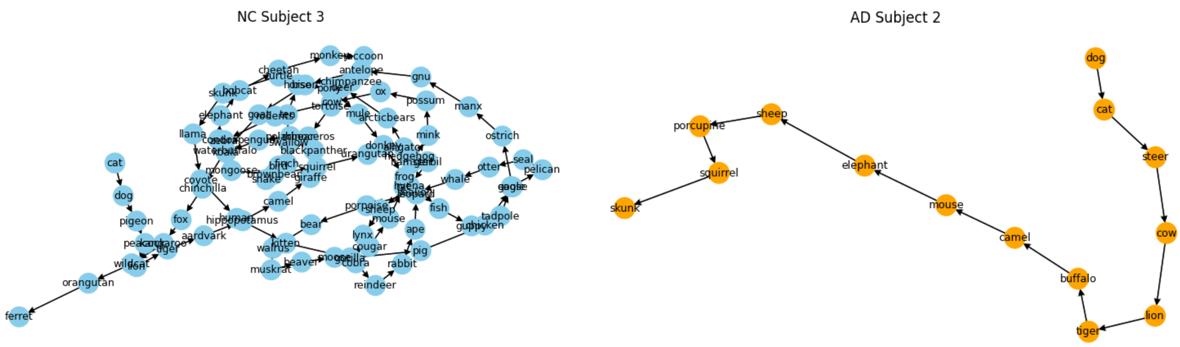


Figure 4. Comparison of NC and AD transition networks.

Subcategory-level transition network

In the subcategory-level approach, each produced word was first assigned to a predefined animal semantic subcategory (e.g., pets, farm animals, wild animals). Networks were then constructed by collapsing the fluency sequence into a sequence of subcategory labels. In this network, nodes represent semantic subcategories rather than individual words, and directed edges encode transitions between subcategories in the order they were retrieved. Repeated transitions within the same subcategory were retained, enabling the analysis of perseveration within local semantic regions as well as failures to transition between distant subcategories. This higher-level representation emphasizes global navigation patterns across semantic space.

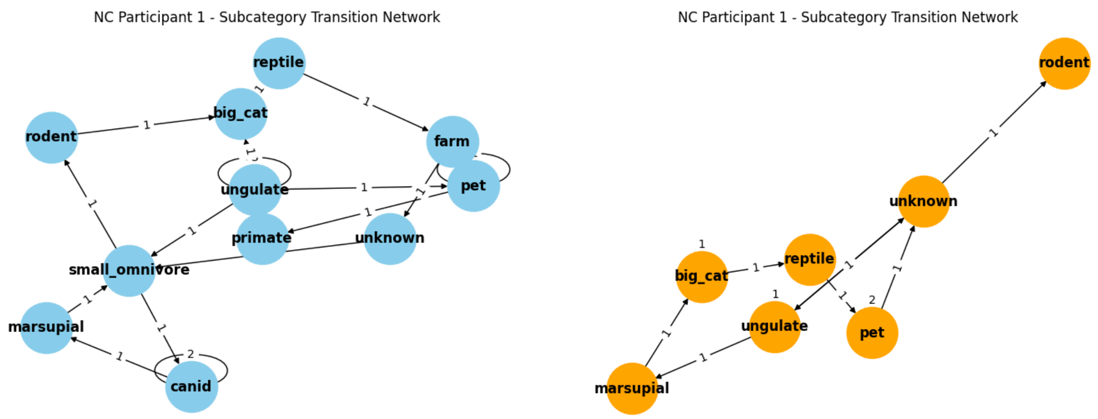


Figure 5. Comparison of NC and AD subcategory transition networks.

Together, these two network constructions allow retrieval dynamics to be analyzed at complementary levels: fine-grained word-to-word transitions and higher-level movements between semantic subcategories.

2.7.2. Network Metrics characterization

To characterize the structure and retrieval dynamics of the transition networks, a set of graph-theoretic metrics was computed at both the word-level and subcategory-level transition networks. Metrics were selected to capture complementary aspects of semantic search that are known to be altered in Alzheimer's disease, including reduced exploration, increased local recurrence, and impaired transitions across semantic regions.

In Word-level transition networks

At the word level, metrics were designed to quantify lexical access, local repetition, and sequential diversity during retrieval. The number of nodes corresponds to the total number of unique words produced, providing a basic measure of accessible vocabulary. The number of edges reflects the total number of transitions between successive words, including repeated transitions. Average degree and network density were computed to quantify overall connectivity, with higher density indicating more repetitive or constrained retrieval patterns.

Impairments in retrieval control were specifically assessed through measures of local recurrence. The self-loop rate, defined as the proportion of immediate word repetitions, directly reflects perseverative behavior and difficulty disengaging from a retrieved item. Complementarily, perseverations and the number of repeated nodes quantify repeated word production across the trial, both of which are classic behavioral markers of executive and retrieval deficits in AD. To capture the ability to sustain novel retrieval sequences, the longest unique path was computed as the maximum number of consecutive words produced without repetition.

In addition, intrusive words, defined as items not belonging to the target semantic category, were identified using WordNet (a lexical database for English) [12] and counted as indicators of impaired semantic control or attentional lapses.

Community detection was not applied at the word level, as these networks are inherently driven by temporal order. Their structure is predominantly sequential, with occasional loops, rather than modular, making the concept of densely connected communities less meaningful in this context.

Subcategory-level transition network

At the subcategory level, networks were analyzed to capture higher-order semantic navigation patterns. The number of nodes represents the total number of distinct semantic subcategories accessed during the trial, while the number of edges corresponds to the total number of transitions between subcategories. Average and median degree, computed from weighted edge counts, were used to summarize transition richness and balance.

Global navigational efficiency was assessed using network density, with higher density suggesting restricted movement among a limited set of subcategories. The self-loop rate at the subcategory level quantifies prolonged persistence within a single semantic region and is hypothesized to be elevated in AD due to reduced switching ability.

Measures of exploratory behavior included the number of repeated subcategories, the longest unique path without revisiting a subcategory, and the jump count, defined as transitions between different subcategories. Reduced jump counts and shorter unique paths are expected to reflect impaired long-range transitions and limited semantic foraging in AD.

To enable fair comparisons across participants with varying fluency output, all metrics were normalized by the total number of nodes in the corresponding network.

2.7.3. Feature Preparation and Normalization

For each participant, the metrics extracted from the word-level and subcategory-level transition networks were aggregated into feature vectors representing retrieval dynamics at complementary semantic scales. Metrics derived from the two network types were treated as separate feature sets and analyzed independently, as well as in combination, to assess their relative and joint discriminative power.

To account for inter-subject variability in fluency output and network size, all metrics, in exception those intrinsically defined as counts of unique elements (e.g., number of nodes or number of subcategories), were normalized by the total number of nodes in the corresponding network.

2.7.4. Classification and Statistical Analysis

To assess the discriminative power of transition network metrics, classification analyses were conducted to distinguish AD patients from NC. Feature sets were derived from word-level networks, subcategory-level networks, and their combination, allowing evaluation of their individual and joint contributions to classification performance.

A supervised machine learning approach was employed, using each participant's feature vector of normalized network metrics as input and the diagnostic label (AD or NC) as the target variable. Random Forest and Logistic Regression classifiers were trained and evaluated. Model performance was assessed using leave-one-out cross-validation (LOOCV), which maximizes training data usage while providing an unbiased estimate of generalization performance in this relatively small dataset. Metrics including accuracy, sensitivity, specificity, and the confusion matrix were computed to quantify classification performance.

To identify which network features contributed most to discrimination, feature importance was extracted. Comparisons between word-level and subcategory-level networks, as well as their combination, were used to determine whether integrating metrics from both levels improved the ability to distinguish AD from NC participants.

3. Results

3.1. Global Topological Properties of Semantic Networks

Firstly, to validate the topological nature of the constructed networks, we compared the global metrics of the real networks against their randomized (shuffled) peers. Importantly, both AD and NC networks proved to be significantly different from a random graphs ($p < 0.001$ for most metrics), confirming that the semantic acquisition process follows a non-random structural organisation. Specifically, real networks showed lower average degree and density than their random counterparts. Moreover, the lengths of the actual shortest paths ($\hat{l}$) were significantly longer than those expected by random chance (e.g., in NC: $\hat{l}_{real} = 3.94$ vs $\hat{l}_{shuffled} = 3.31$, $p < 0.001$). The observed differences reflect a reduction in network scale and navigational range, rather than genuine differences in processing efficiency.

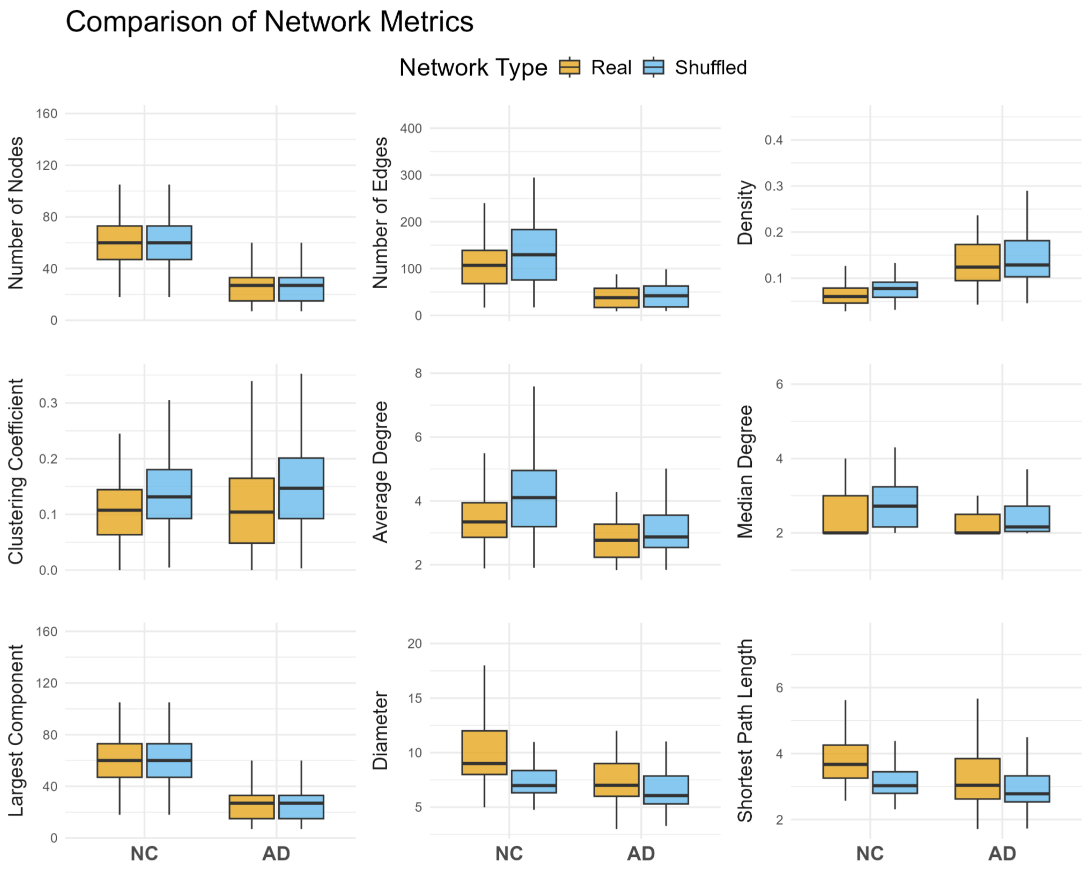


Figure 6. Comparison of network metrics between NC and AD Groups and Type of Network. NC: Normal Control, AD: Alzheimer's Disease

The analysis of global network metrics revealed profound structural differences between the semantic networks of AD patients and NC.

The most prominent difference was observed in the network size. The AD group produced significantly fewer nodes ($N = 26.44 \pm 12.96$) compared to the NC group ($N = 61.07 \pm 26.60$, $p < 0.001$). Consistent with this reduction, the number of edges was also drastically lower in patients (39.79 ± 23.95) than in controls (108.57 ± 61.09, $p < 0.001$). Furthermore, the average degree (k) was significantly reduced in the AD group (2.85 ± 0.80) compared to controls (3.36 ± 0.83, $p < 0.001$), indicating a decline in connectivity, with words having fewer associations on average.

The network density (ρ) shows an inverse pattern, being significantly higher in the AD group (0.14 ± 0.06) compared to the NC group (0.06 ± 0.02, $p < 0.001$). This increase should be interpreted with caution, as it is likely an artifact driven by the drastic reduction in the network size (N) rather than an indicator of enhanced integration, as density scales inversely with the number of nodes in sparse networks.

Regarding network dimensions, the AD networks showed a significantly smaller diameter (7.79 ± 3.14) compared to NC networks (10.02 ± 3.45, $p < 0.001$). While shorter paths usually suggest higher efficiency, in this context, they reflect the structural fragmentation and the smaller size of the AD semantic space.

Finally, no significant differences were found in the global clustering coefficient between groups (0.12 ± 0.10 for AD vs. 0.11 ± 0.06 for NC, $p = 0.269$). This suggests that the local tendency of words to form categorical triangles remains relatively preserved despite the massive loss of lexical content and global connectivity.

3.2. Small-World Organization in AD and NC Networks

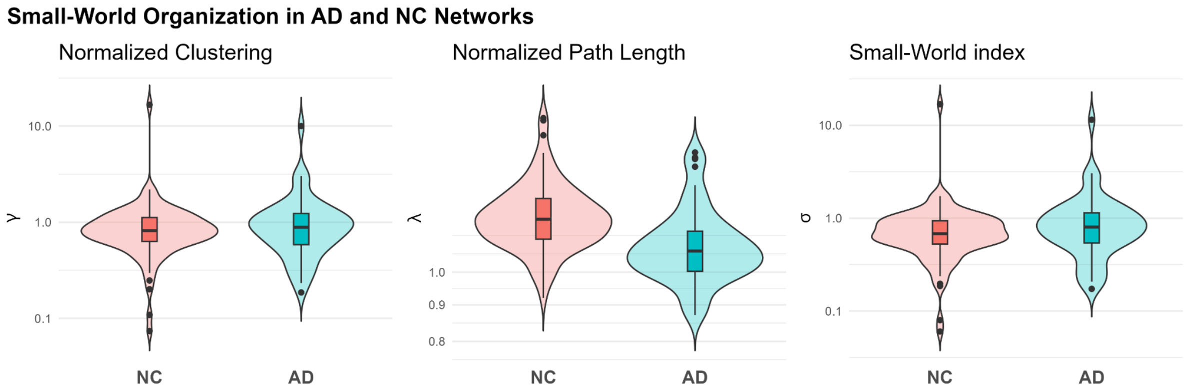


***Figure 7.** Small-World organization among AD and NC semantic networks.*
NC: Normal Control, AD: Alzheimer's Disease

To determine if the semantic networks exhibit a small-world topology -characterized by high local clustering and short global path lengths relative to random networks- we compared the normalized clustering coefficient (γ), normalized path length (λ), and the small-world index (σ) between the AD and NC groups.

First, regarding the Normalized Clustering (γ), no significant differences were observed between the groups ($p = 0.97$). Both groups exhibited median values close to 1.0 (NC: 0.96 ± 1.67; AD: 0.96 ± 1.36). This indicates that, in both healthy and pathological networks, the level of local clustering is practically identical to what would be expected in a random network of the same size. The overlap between the distributions suggests that the disease does not differentially affect local clustering relative to the random baseline.

In contrast, the Normalized Path Length (λ) showed a statistically significant reduction in the AD group compared to controls ($p < 0.001$). The NC group presented a mean of 1.20 ± 0.13, whereas the AD group had a mean of 1.09 ± 0.13. This suggests that the semantic paths in AD networks are closer to the randomness baseline than those in NC networks, likely due to the drastically reduced network size.

Finally, the Small-World Index (σ), displayed in the third panel, revealed no significant differences between the groups ($p = 0.75$). The mean values were 0.84 ± 1.69 for NC and 0.92 ± 1.53 for AD. Moreover, the distributions for both groups fall largely below the threshold of $\sigma > 1$. This finding implies that neither the AD nor the NC semantic networks qualify as "small-world" networks. The topology in both cases is dominated by the lack of high clustering ($\gamma \approx 1$), resulting in an organisation that looks more like a random network than a small-world architecture.

3.3. Community Structure and Modular Organization

The statistical comparison between real and shuffled networks reveals that semantic structure is significantly non-random across multiple metrics ($p < 0.001$). Both the Louvain and Infomap algorithms confirm that the real networks possess a significantly higher degree of modular organization than expected by random chance.

The modularity (Q) of the real networks was found to be 0.48 ± 0.09, which is significantly higher than that of the shuffled networks (0.41 ± 0.09, $p < 0.001$). Additionally, the real networks tended to be decomposed into a slightly higher number of communities (6.09 ± 1.82) compared to the random models (5.86 ± 1.82, $p < 0.001$), with a correspondingly smaller average community size (7.31 vs 7.62 nodes).

This trend was further corroborated by the Infomap algorithm, where the real networks exhibited a higher modularity ($Q = 0.45 \pm 0.14$) compared to the randomized counterparts (0.35 ± 0.14). Infomap also detected a significantly higher number of communities in the real networks (7.39 ± 3.89) versus the shuffled ones (6.50 ± 3.89, $p < 0.001$).

In contrast to the modularity-based and flow-based methods, the Degree-Corrected Stochastic Block Model (DCSBM) showed less sensitivity to the specific structure of these semantic networks. In terms of the number of communities, no significant differences was found, with the model identifying a single community ($N_c = 1$) for both real and shuffled networks in the majority of cases. However, the Description Length (DL) was significantly lower for the real networks (313.58 ± 254.76) compared to the shuffled ones $(344.93 \pm 254.76$, $p < 0.001)$. This reduction in DL indicates that, despite the lack of detected communities by this specific algorithm, the real networks contain structural patterns.

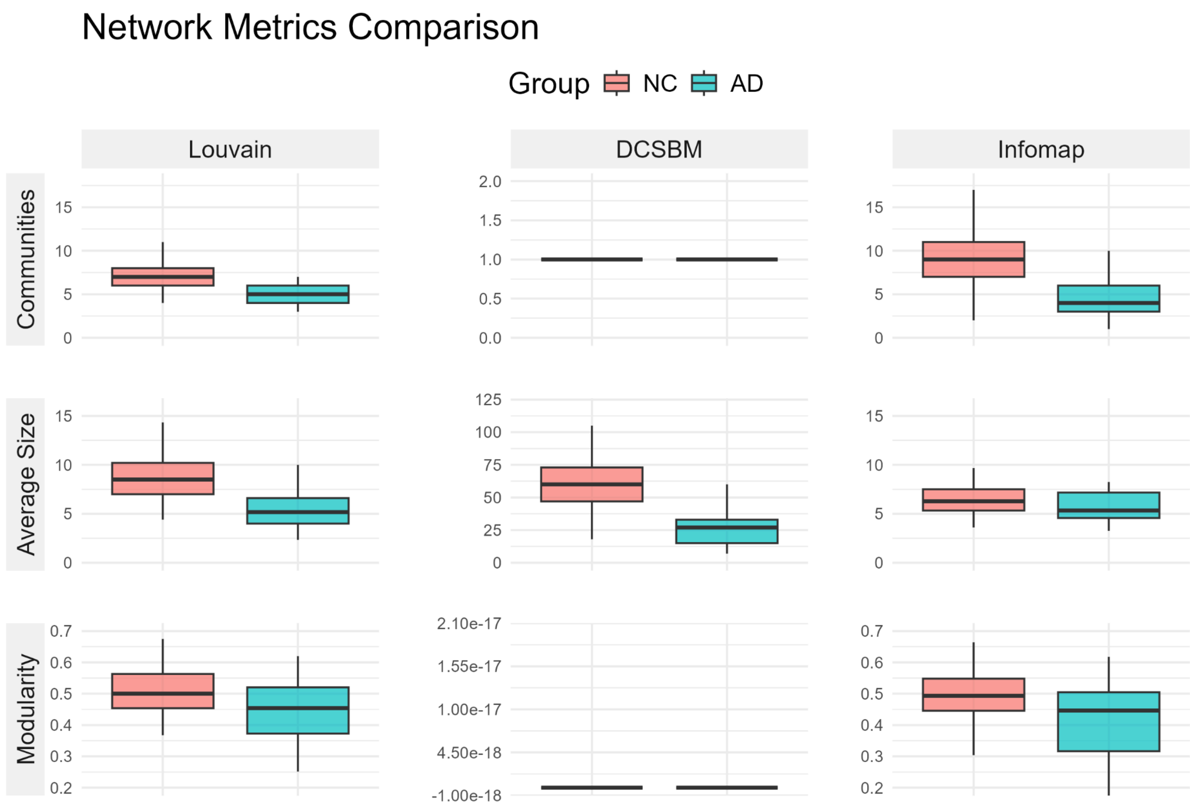


Figure 8. *Normal Control ((NC) vs Alzheimer's Disease (AD) Global metrics* comparison.

When analyzing the real networks among NC and AD groups, Louvain and Infomap successfully detected structures in the semantic networks. Both revealed that the semantic networks of healthy subjects are significantly more fragmented than those of AD patients. The NC group exhibited a higher number of communities (Louvain: 6.92 ± 1.66; Infomap: 9.24 ± 3.51) compared to the AD group (Louvain: 4.79 ± 1.20; Infomap: 4.44 ± 2.35). These differences are statistically significant $(p < 0.001)$ and indicate that healthy semantic memory maintains a richer variety of distinct animal sub-categories.

Louvain detected significantly larger communities (average community size) in the NC group $(8.56 \pm 2.43$ nodes) compared to the AD group $(5.32 \pm 1.76, p < 0.001)$. However, the Infomap algorithm found no significant difference in the average community size $(p = 0.89$) between groups $(NC = 6.94,\ AD = 7.04)$. This suggests that while the number of semantic categories decreases in Alzheimer's disease, the typical size of the remaining categories (detected by flow-based methods like Infomap) remains relatively stable.

It is important to highlight that the Degree-Corrected Stochastic Block Model (DCSBM) did not detect significant modular fragmentation as it consistently identifies a single community across all subjects. This suggests that while the networks exhibit high modularity, there is no evidence of a block structure, probably due to the small size and high density of the semantic networks.

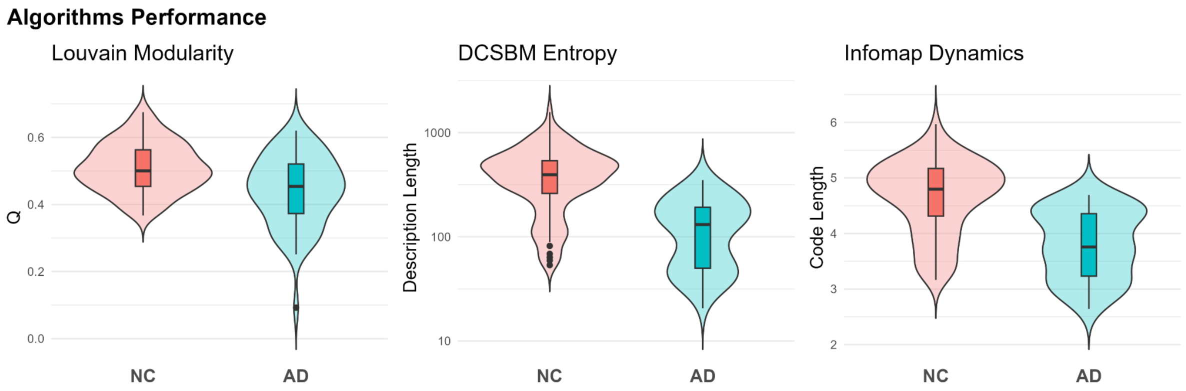


Figure 9. *Normal Control ((NC) vs Alzheimer's Disease (AD) Algorithms performance.*

The second analysis evaluated the robustness and information content of the network structure using the quality functions of each algorithm.

For the Louvain algorithm, modularity was used as a quality metric. The clarity of the network structure is significantly degraded in the AD group. The NC group presented higher modularity values (0.51 ± 0.07) compared to the AD group (0.44 ± 0.10, $p < 0.001$). This higher modularity value in healthy subjects implies well-defined semantic categories, whereas in AD subjects, these boundaries become more diffuse.

The complexity of the networks was assessed using DCSBM's Description Length, which was statistically higher in NC group (426.84 ± 260.10) than in AD group (133.48 ± 88.70, $p < 0.001$).

Also, the Code Length (CL) was used to evaluate Infomap's performance. it was also statistically higher in NC group (4.66 ± 0.66) than in AD group (3.80 ± 0.61, $p < 0.001$).

3.4. Subject Similarity Network

The Subject Similarity Network (SSN) analysis aimed to determine if AD and NC subjects could be distinguished by the topology of their semantic networks, without supervised training. We evaluated the partition quality of five different algorithms against the true clinical diagnoses.

Algorithm	Ground Truth	Louvain	DCSBM	Infomap	Spectral Clustering	Agglomerative Clustering
Number of Communities	2	9	8	12	2	2
True Negatives	97	82	84	82	82	28
True Positives	61	48	42	46	46	55
Accuracy	1	0.82	0.80	0.81	0.78	0.65

Table 1. *Algorithms comparison when classifying subjects from Subjects Similarity Network.*

As shown in Table 1, the ability of the SSN to recover the ground truth varies by algorithm.

The results of Louvain, DCSBM and Infomap algorithms reveal that the topological structure of semantic memory does not simply divide into two homogenous clusters (AD vs NC) but rather exhibits a more complex organization. That is why Spectral and Agglomerative Clustering algorithms were contemplated.

Louvain achieved the highest overall accuracy and the best balance between sensitivity and specificity, identifying 9 communities.

DCSBM obtained the highest number of True Negatives, subjects who truly belong to the NC group. This suggests that it is particularly effective at characterizing the healthy topological profile.

Infomap produced the most fragmented partition (12 communities), consistent with its flow-based approach detecting finer local structures. However, it maintains good accuracy.

Finally, binary partition algorithms such as Spectral and Agglomerative Clustering had the worst accuracy. This indicates that forcing a binary segmentation has some limitations. Specifically, Agglomerative Clustering had some troubles by identifying subjects from the NC group (low True Negatives).

The algorithms that achieved the highest classification accuracy did not identify two communities. This over-segmentation suggests that the clinical categories of NC and AD are topologically heterogeneous. These algorithms likely detect sub-phenotypes by separating subjects by disease severity (e.g., mild vs severe AD).

3.5. Network Reorganization and Exploratory Structural Analysis

3.5.1. Interpretation of Transition Network Metrics

For each fluency trial, transition networks were constructed and a set of graph-theoretic and behavioral metrics was extracted, capturing lexical diversity, connectivity, and repetition-related phenomena.

Word-level transition networks

Analysis of the word-level networks revealed marked differences in network size between NC and Alzheimer's disease AD participants. NC participants produced substantially larger networks, with a higher number of unique words and transitions (mean ≈ 61) compared to AD participants (mean ≈ 26), reflecting reduced lexical access in AD. In contrast, average and median degree were similar across groups, indicating that local connectivity and immediate sequential transitions were largely preserved in both populations.

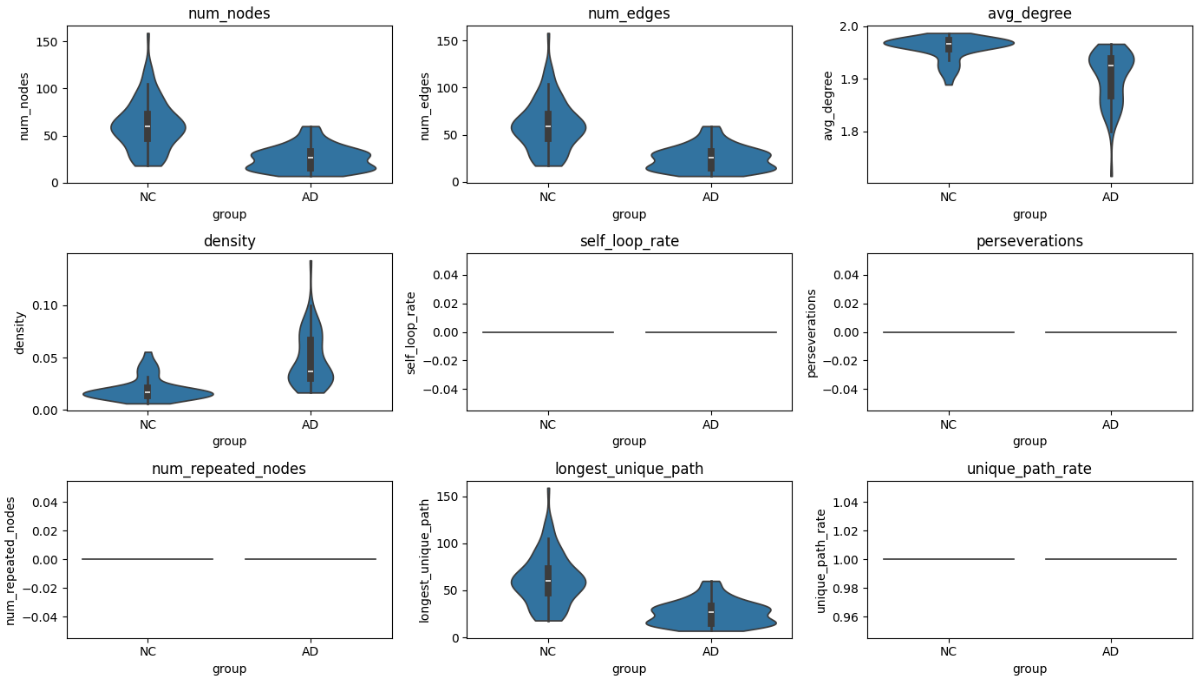


Figure 10. *Graph-metrics comparison among NC and AD networks*

Network density was higher in AD networks, which appears to be a size-driven effect rather than an indicator of increased connectivity: smaller networks inherently produce higher density when edges are maintained. Surprisingly, no repeated nodes or short cycles were observed in either group, resulting in identical repetition rates and unique path lengths. The longest unique path closely matched the total number of nodes, suggesting that differences in fluency performance were primarily driven by the total number of retrieved items rather than local looping or alterations in word-order dynamics.

Prior literature had suggested that AD patients would show higher rates of repeated words and self-loops, reflecting perseverative retrieval [13]. The absence of such repetitions in this dataset suggests that overt perseveration was either rare under the current testing conditions, or that previous assumptions overestimated its prevalence. Overall, at the word level, group differences were largely expressed through network size, with minimal variation in local sequential structure.

Subcategory-level transition networks

In contrast, the analysis of subcategory-level networks revealed meaningful differences in how participants navigated semantic space. NC participants produced larger networks with more transitions and higher jump counts, suggesting a greater capacity to flexibly explore semantic subcategories and generate diverse sequences. AD participants, by comparison, exhibited smaller networks with higher relative density, indicative of a more constrained exploration pattern. Self-loops and repeated transitions at the subcategory level reflected perseverative tendencies or difficulties in switching between semantic regions, consistent with expected deficits in controlled retrieval and strategic search.

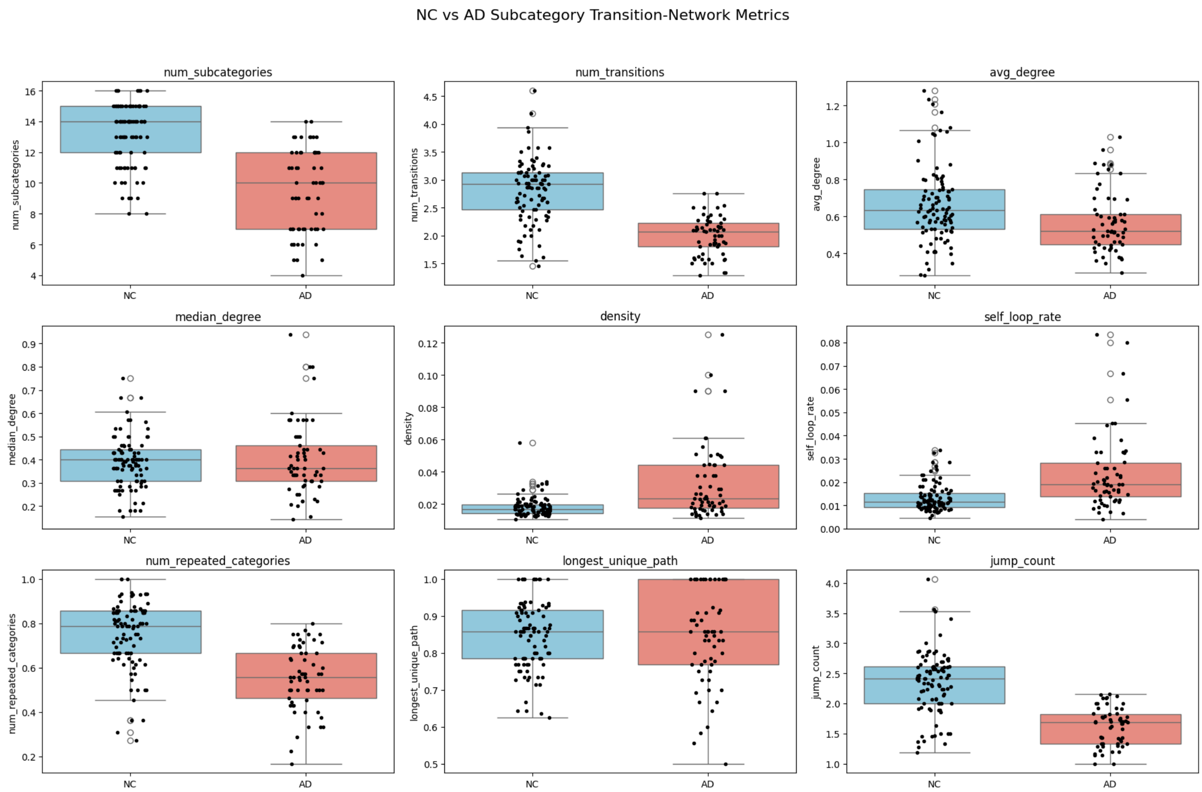


Figure 11. *Graph-metrics comparison among NC and AD networks*

Despite these differences, the normalized longest path was similar between groups, indicating that AD participants retained the ability to perform consecutive transitions without repetition. However, overall variability and complexity in subcategory navigation were reduced. These patterns highlight a distinction between the two groups: NC subjects show rich, adaptive subcategory exploration, whereas AD subjects demonstrate restricted, repetitive, and less flexible patterns, consistent with impairments in task organization, planning, and semantic search.

3.5.2. Classification of AD vs. NC Using Network Metrics

To evaluate the predictive value of transition network features, supervised classification analyses were performed using metrics extracted from both word-level and subcategory-level networks. Separate models were trained for each feature set, as well as for the combined feature set, to assess the relative and joint contribution of word-level and subcategory-level retrieval dynamics.

Word-Level Network Classification Results

Classification analyses using word-level fluency network metrics were performed to distinguish AD patients from normal controls NC. Using Random Forest, the model achieved an accuracy of 72%, with a precision, recall, and F1-score of 0.73, 0.74, and 0.72, respectively.

Logistic Regression achieved higher accuracy (81%) and perfect sensitivity for AD, correctly identifying all AD participants, while Random Forest showed slightly lower performance (72% accuracy) and a tendency to misclassify NC participants as AD. These results indicate that group differences are largely captured by linearly separable features, primarily reflecting reduced network size and retrieval diversity in AD rather than complex nonlinear interaction patterns.

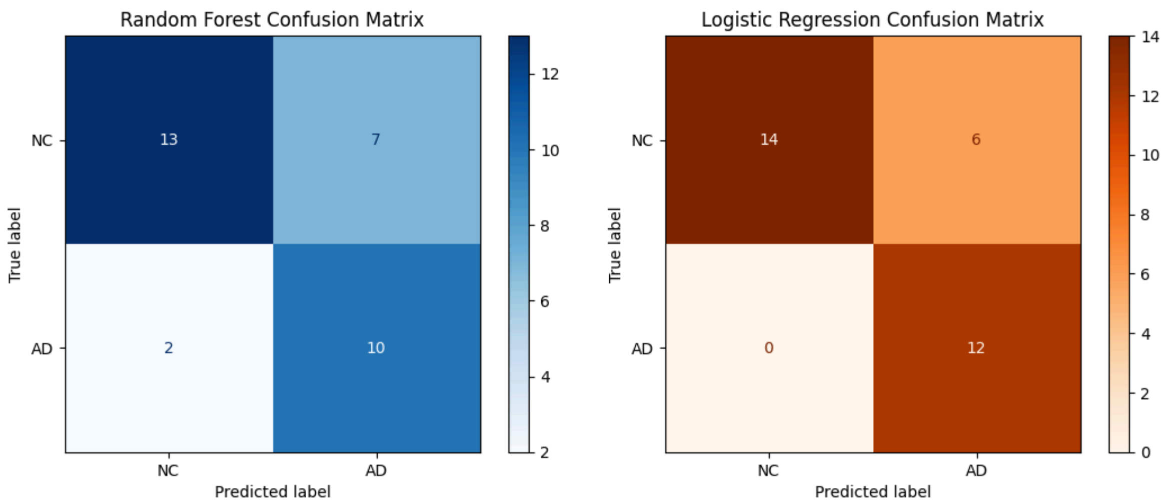


***Figure 12.** Graph-metrics comparison among NC and AD Word-Level networks.*

Using leave-one-out cross-validation (LOOCV), which maximizes training data usage while providing an unbiased estimate of generalization performance, the combined word-level metrics achieved an accuracy of 75.3%, confirming the robustness of the findings despite the relatively small sample size.

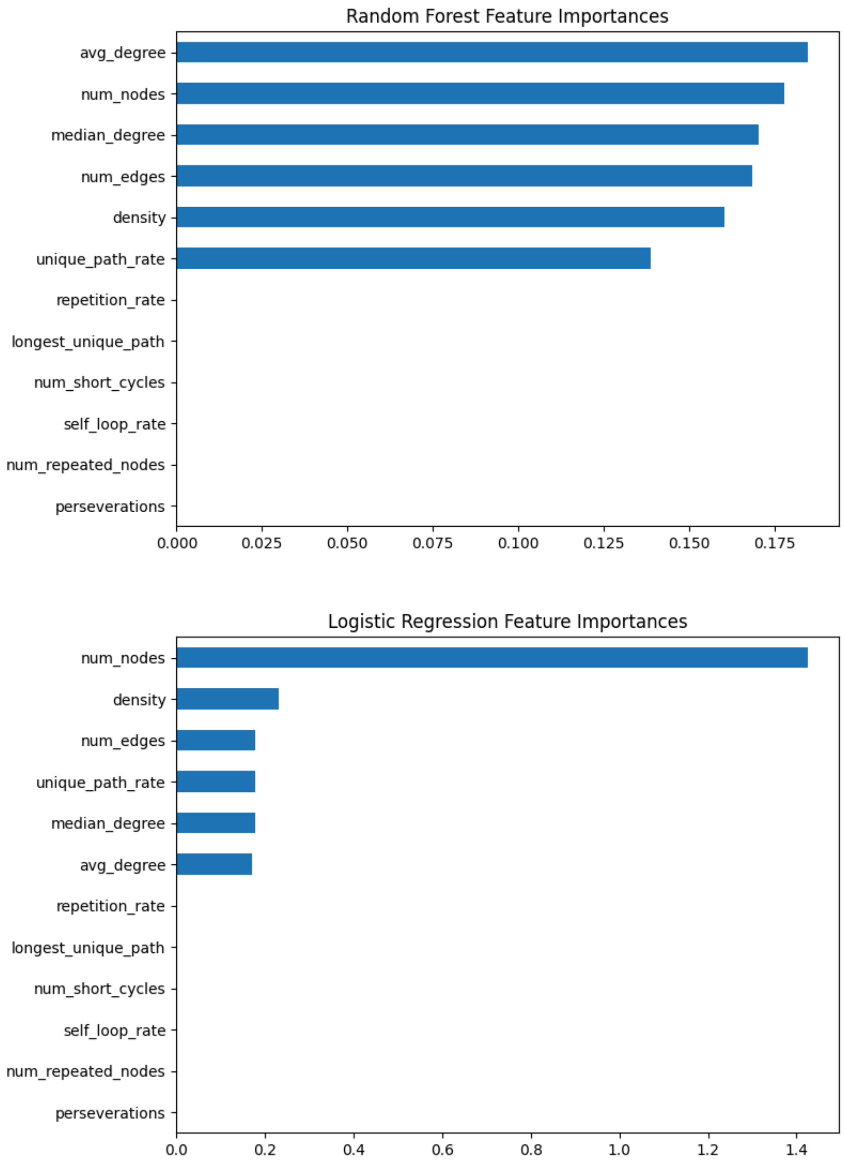


Figure 13. *Feature importance from World-Level Network classification.*

Feature importance analysis of the Logistic Regression model revealed that classification was primarily driven by global network size and diversity metrics. Number of nodes, number of edges, and longest unique path showed the highest coefficients, indicating that reduced vocabulary access and shortened unique retrieval sequences were the strongest discriminative features between NC and AD participants. Metrics related to repetition, such as self-loop rate, perseverations, and repeated nodes, showed zero contribution, consistent with the absence of overt repetitions in the data. These findings suggest that, in this dataset, AD-related impairment is characterized more by early exhaustion of semantic retrieval rather than by increased perseverative behavior.

Subcategory-Level Network Classification Results

Using subcategory-level network metrics, Random Forest classified AD and NC participants with 81% accuracy, correctly identifying 16 of 20 NC and 8 of 12 AD participants. Logistic Regression achieved 75% accuracy, showing slightly lower performance.

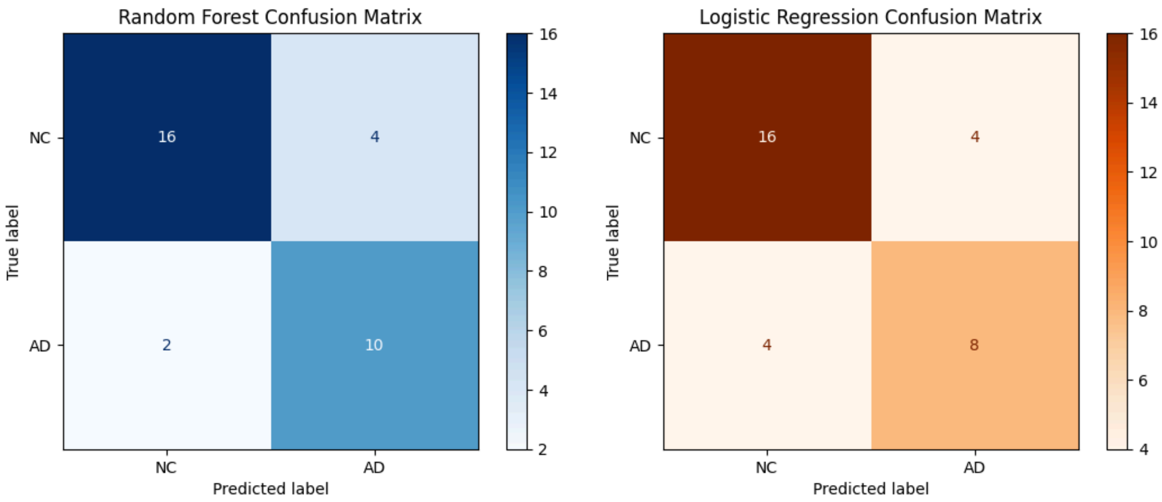
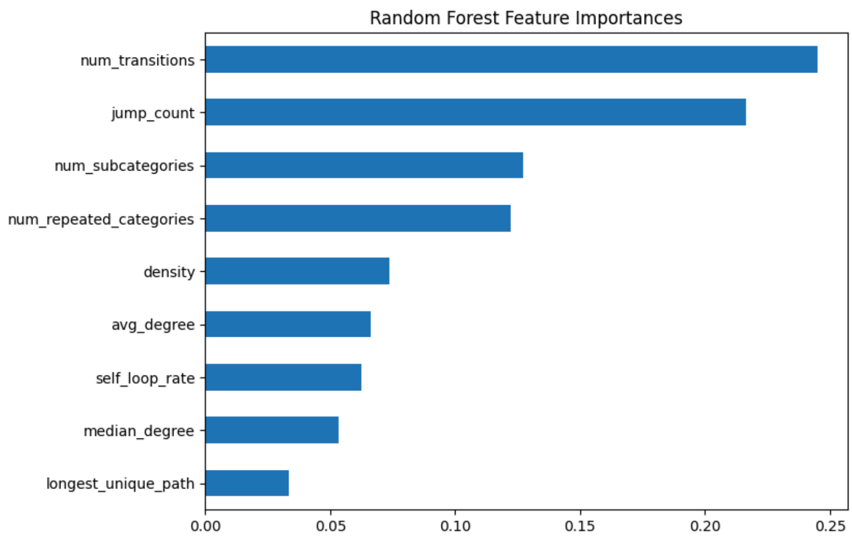


Figure 14. Graph-metrics comparison among NC and AD Subcategory-Level nnetworks.

With LOOCV, the subcategory-level metrics reached an accuracy of 87.3%, confirming that transitions between semantic subcategories are highly informative for distinguishing AD from NC participants.



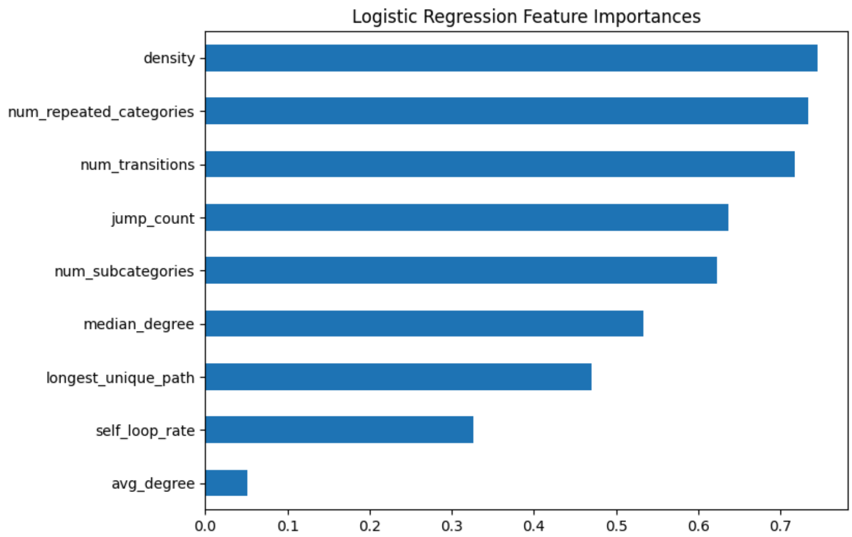


Figure 15. *Feature importance from Subcategory-Level Network classification.*

Feature importance analysis showed that the most informative metrics were number of transitions, jump count, number of subcategories, and network density. Other features, such as node degree or longest unique paths, were less important. These results suggest that AD-related impairment primarily affects the ability to flexibly navigate between semantic subcategories, complementing the word-level findings.

Combined Word-Level and Subcategory-Level Network Classification Results

Using combined word-level and subcategory-level network metrics, Random Forest achieved 84% accuracy, correctly classifying 86 of 97 NC participants and 47 of 61 AD participants. Logistic Regression reached 80% accuracy, with slightly lower recall for AD participants. These results show that combining metrics from both levels improves discrimination compared to using either feature set alone.

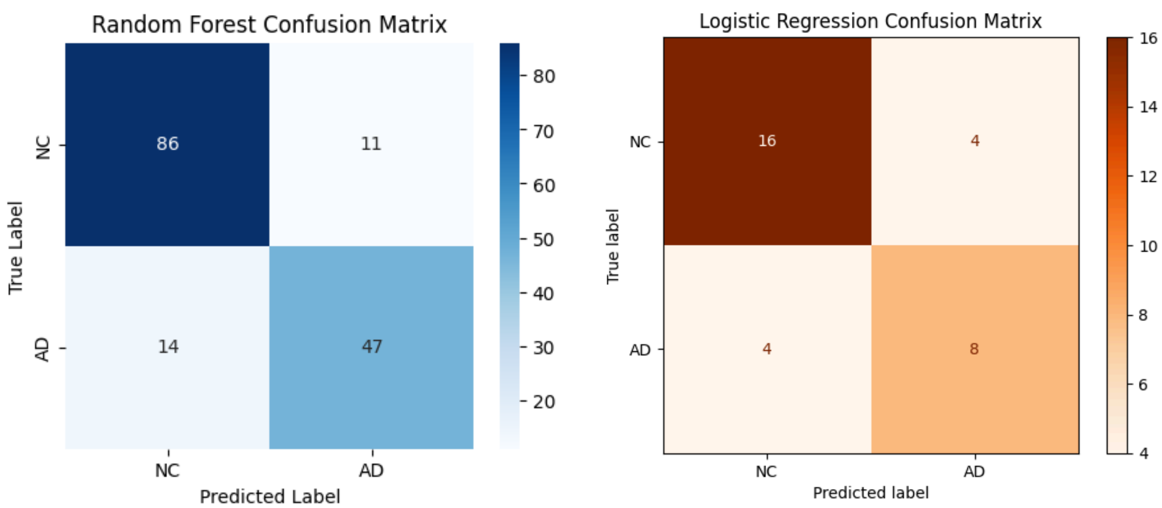
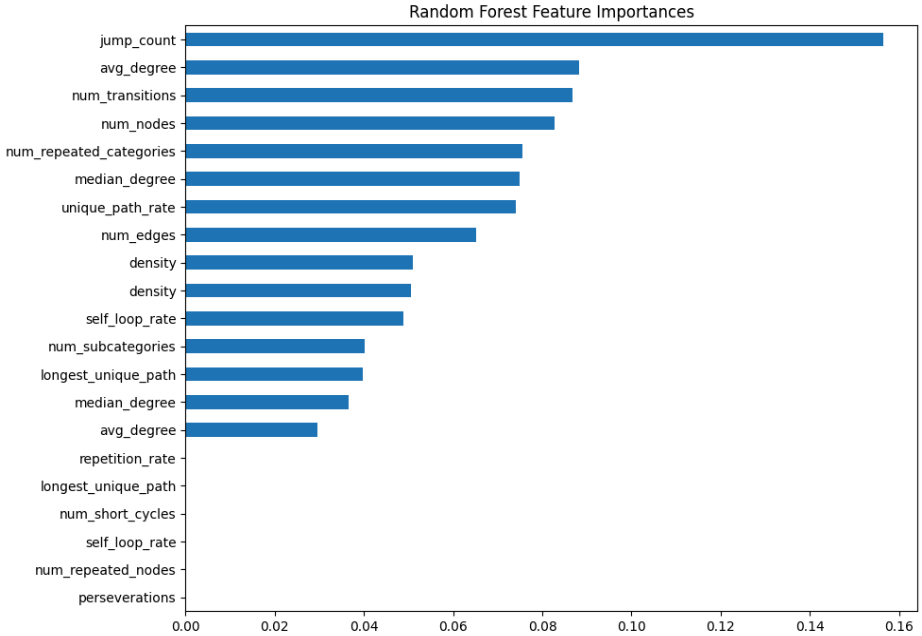


Figure 16. *Graph-metrics comparison among NC and AD combined networks*

Feature importance analysis indicated that the most informative metrics were network size and connectivity (number of nodes, edges, degree, density, and unique path rate) and subcategory transitions (number of subcategories, number of transitions, jump count, and repeated categories). Metrics related to repetition or short cycles had minimal impact.



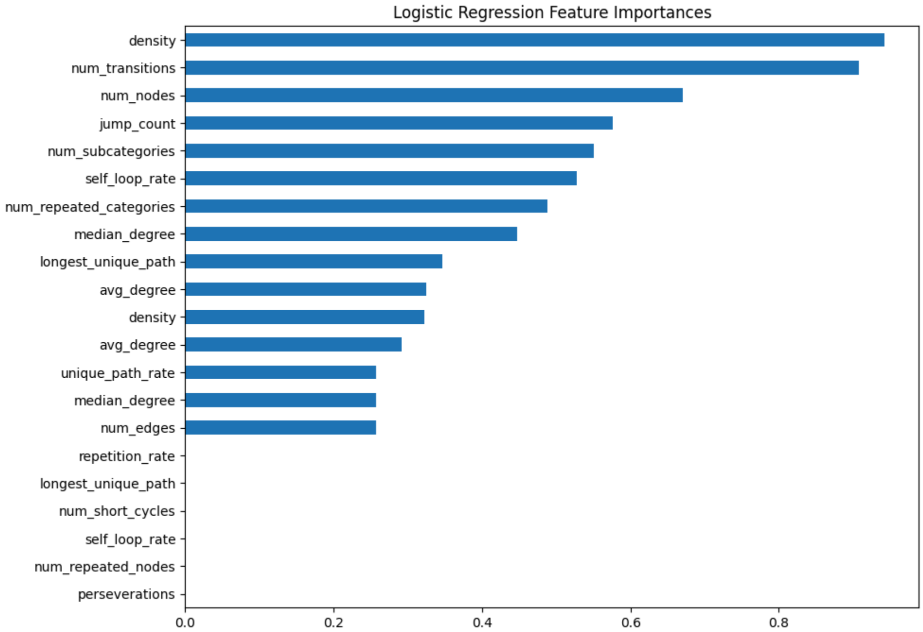


Figure 17. *Feature importance from Combined Network classification.*

Overall, the findings suggest that AD-related impairment is characterized by reduced network size, lower transition diversity, and less flexible navigation of semantic categories, and that combining word-level and subcategory-level features provides the strongest discrimination between NC and AD participants.

4. Discussion

The present study aimed to reproduce and extend the work of *Zemla & Austerweil (2019)* by investigating the topological, modular, and navigational properties of semantic networks in subjects with Alzheimer's Disease (AD) and Normal Cognitive controls (NC). Our results not only corroborate the structural deterioration associated with dementia but also offer new perspectives on how lexical retrieval dynamics (navigation) may serve as a more sensitive marker than static network structure.

4.1. Network Size and Density artifact

The most consistent finding in our global metric analysis is the drastic reduction in network size within the AD group ($N \approx 26$) compared to the NC group ($N \approx 61$). This lexical poverty leads to a chain reaction in other topological metrics that require careful interpretation. Contrary to expectations, we found that AD networks have a density (ρ) and smaller diameter. However, we argue that this does not reflect better semantic integration, but rather a mathematical artefact derived from network size. In very small graphs, the probability of any two nodes being connected increases artificially as the search space decreases. Therefore, AD networks are not more efficient, but they are topologically collapsed: the semantic space has contracted so severely that the remaining concepts appear closer to one another simply because there are fewer intermediate concepts.

4.2. Absence of Small-Worldness

Unlike previous studies reporting small-world properties in semantic networks, our results showed a small-world index of $\sigma \approx 1$ for both AD and NC groups. This may suggest that, under the constraints of the animal fluency task, the generated networks do not significantly differ from random graphs in terms of local clustering. However, this conclusion must be interpreted in the context of the null model construction. The null model in this study did not generate random networks from scratch using Erdős-Rényi model (which assumes uniform edge probability). Instead, they were obtained by randomly permuting the order of items within the original fluency lists. This method preserves the number of nodes and, more importantly, the frequency of occurrence of elements. Small-world topology would have been detected if we had compared our empirical networks to a completely random null model.

4.3. Dissolution of Modular Structure

Community analysis revealed that AD deterioration is not uniform, affecting categorical organisation. Modularity was significantly lower in AD patients ($Q = 0.44$) than in NC subjects ($Q = 0.51$), indicating that the boundaries between semantic subcategories (e.g., farm animals, domestic animals,...) become diffuse. Our Louvain and Infomap results showed that AD patients recover fewer, smaller communities, supporting the hypothesis that AD destroys the clustered structure of semantic memory, resulting in disorganised searching.

4.4. Heterogeneity in Subject Similarity Network

The Subject Similarity Network (SSN) analysis offered an interesting unsupervised validation. Although we expected two clear communities (AD vs. NC), no algorithm achieved perfect separation. The modularity-based (Louvain), probabilistic (DCSBM), and flow-based (Infomap) methods successfully recovered the diagnostic labels with ≥ 80 accuracy (up to **82% accuracy** with Louvain). This confirms that topological metrics contain a strong diagnostic signal, provided that the analysis allows for the natural heterogeneity of the population (i.e., multiple communities) rather than enforcing a strict binary separation. Despite achieving the highest classification accuracy, they identified more than two communities. Far from being an error, this over-segmentation suggests that the clinical categories of NC and AD are topologically heterogeneous. These algorithms likely detect sub-phenotypes by separating subjects by disease severity (e.g., mild vs severe AD).

4.5. Semantic Navigation: Failures in Switching

The extension of this study into transition networks (world-level and subcategory-level) provided the highest discriminative power. While static metrics showed some similarities, dynamic metrics revealed clear functional deficits.

At the subcategory level, AD patients showed a significant reduction in jump count (transitions between categories) and higher transition density. This aligns with literature on executive switching deficits: patients become stuck in local patterns of thinking (implicit perseveration) or make random transitions without a global search strategy. The fact that the combined classification (word + subcategory level) achieved approximately **84% accuracy** underscores that AD pathology is both a loss of storage (nodes) and a failure in the access mechanism (transition edges).

4.6. Limitations

It has to be mentioned that it represents a limitation that algorithms such as Louvain, DCSBM, Infomap, and Spectral Clustering are stochastic, meaning that may yield slightly different partition results in each iteration due to random initialization or node processing order.

Finally, it is possible that more abstract domains different from animals would show different patterns of deterioration. Also, the lack of longitudinal data prevents us from confirming whether the loss of modularity precedes the collapse of network size, which would be a key for early diagnosis.

4.7. Conclusions

In conclusion, this study demonstrates that Alzheimer's Disease dismantles semantic memory by reducing its size, blurring its modular structure, and limiting the flexibility of navigation between concepts. The integration of dynamic transition metrics was more effective than using static topological metrics in patient classification, suggesting that the way in which we search our memory is as important as, if not more important than, the information we retrieve.

References

[1] Zemla, J. C., & Austerweil, J. L. (2019). Analyzing Knowledge Retrieval Impairments Associated with Alzheimer's Disease Using Network Analyses. Complexity, 2019, 4203158. https://doi.org/10.1155/2019/4203158

[2] T. Vos, C. Allen, M. Arora, M. Barber R, A. Bhutta Z, and A. Brown, "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015," The Lancet, vol. 388, no. 10053, pp. 1545– 1602, 2016.

[3] F. J. Huff, S. Corkin, and J. H. Growdon, "Semantic impairment and anomia in Alzheimer's disease," Brain and Language, vol. 28, no. 2, pp. 235–249, 1986.

[4] A. M. Collins and M. R. Quillian, "Retrieval time from semantic memory," Journal of Verbal Learning and Verbal Behavior, vol. 8, no. 2, pp. 240–247, 1969.

[5] Bassett DS, Sporns O. Small-world human brain networks: Perspectives and challenges. Neuroscience & Biobehavioral Reviews. 2017 Jun;77:286-300. doi:10.1016/j.neubiorev.2017.03.018.

[6] R. F. I. Cancho and R. V. Sol´e, "The small world of human language," Proceedings of the Royal Society B Biological Science, vol. 268, no. 1482, pp. 2261–2265, 2001.

[7] Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech*. 2008;2008(10):P10008.

[8] Smiljanić J, Blöcker C, Holmgren A, Edler D, Neuman M, Rosvall M. Community detection with the map equation and Infomap: theory and applications [Internet]. arXiv; 2023 [cited 2026 Jan 3]. Available from: https://arxiv.org/abs/2311.04036

[9] Newman MEJ. Modularity and community structure in networks [Internet]. arXiv; 2010 [cited 2026 Jan 3]. Available from: https://arxiv.org/abs/1008.3926

[10] Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

[11] Raoux N, Amieva H, Le Goff M, Auriacombe S, Carcaillon L, Letenneur L, Dartigues JF. Clustering and switching processes in semantic verbal fluency in the course of Alzheimer's disease subjects: results from the PAQUID longitudinal study [Internet]. Cortex. 2008 Oct [cited 2026 Jan 3]; 44(9):1188–1196. Available from: https://pubmed.ncbi.nlm.nih.gov/18761132/

[12] Miller GA. WordNet: a lexical database for English [Internet]. Communications of the ACM. 1995 [cited 2026 Jan 3];38(11):39–41. Available from: https://wordnet.princeton.edu/

[13] Dubey, N., Basu, J., Pandey, D., & Ghosh, A. (2025). Switching and its Impact on Perseveration in a Verbal Fluency Task: a Study in Persons With Alzheimer's Clinical Syndrome. Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists, 40(5), 936–944. Available from: https://doi.org/10.1093/arclin/acaf007