

This is an Accepted Manuscript of an article published by Taylor & Francis in AGING, NEUROPSYCHOLOGY, AND COGNITION on August 2021, available at <https://doi.org/10.1080/13825585.2021.1965951>

Cite: Nevado, A., del Rio, D., Pacios, J., & Maestú, F. (2022). Neuropsychological networks in cognitively healthy older adults and dementia patients. *Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition*, 29(5), 903–927. <https://doi.org/10.1080/13825585.2021.1965951>.

## **Neuropsychological Networks in Cognitively Healthy Older Adults and Dementia Patients**

Angel Nevado<sup>ab\*</sup>, David del Rio<sup>ab</sup>, Javier Pacios-Garcia<sup>ab</sup> and Fernando Maestú<sup>ab</sup>

<sup>a</sup>*Experimental Psychology Department, Complutense University of Madrid, Madrid, Spain;*

<sup>b</sup>*Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain*

Corresponding author: Angel Nevado, PhD. School of Psychology, Complutense University of Madrid, Campus de Somosaguas, 28223 Madrid, Spain. Phone: +34 913943110. Email: [anevado@ucm.es](mailto:anevado@ucm.es). Fax: +34 913943189

Word count: 7365

## Neuropsychological Networks in Cognitively Healthy Older Adults and Dementia Patients

Neuropsychological tests have commonly been used to determine the organization of cognitive functions by identifying latent variables. In contrast, an approach which has seldom been employed is network analysis. We characterize the network structure of a set of representative neuropsychological test scores in cognitively healthy older adults and MCI and dementia patients using network analysis. We employed the neuropsychological battery from the National Alzheimer's Coordinating Center which included healthy controls (n=7623), mild cognitive impairment patients (n=5981) and dementia patients (n=2040), defined according to the Clinical Dementia Rating. The results showed that, according to several network analysis measures, the most central cognitive function is executive function followed by language, attention, and memory. At the test level, the most central test was the Trail Making Test B, which measures cognitive flexibility. Importantly, these results and most other network measures, such as the community organization and graph representation, were similar across the three diagnostic groups. Therefore, network analysis can help to establish a ranking of cognitive functions and tests based on network centrality and suggests that this organization is preserved in dementia. Central nodes might be particularly relevant both from a theoretical and clinical point of view, as they are more associated with other nodes, and their disruption is likely to have a larger effect on the overall network than peripheral nodes. The present analysis may provide a proof of principle for the application of network analysis to cognitive data.

Keywords: neuropsychology, network analysis, cognition, Alzheimer's disease, graph theory

## **Introduction**

There has been a rapid increase in recent years in the use of network analysis, also referred to as graph theory, in the fields of psychopathology (Borsboom & Cramer, 2013; Contreras et al., 2019; McNally, 2016), personality (Costantini et al., 2015; Cramer et al., 2012) and neuroimaging (Rubinov & Sporns, 2010; Sporns, 2018). In contrast, much less work has been devoted to understanding the interdependencies between cognitive functions.

The most common approach to study the statistical relationships between neuropsychological test scores is to employ latent variable analysis. Its goal is to identify the underlying factors causing the observed test scores (Dowling et al., 2010; Miyake, Friedman, et al., 2000; Thurstone, 1947). This method has the advantages of reducing the dimensionality of the analysis and mitigating the task impurity problem, whereby performance variability in any single neuropsychological test is influenced by more cognitive mechanisms than the one purportedly measured (Fournier-Vicente et al., 2008; Miyake, Emerson, et al., 2000). For example, memory tests usually depend on attentional mechanisms as well as on domain specific language or visuospatial skills.

On the other hand, assessing statistical dependencies at the test score level can provide important complementary information. Although a degree of association between tests measuring different cognitive domains is to be expected, this type of analysis can also reveal genuine dependencies between the different cognitive mechanisms, above correlations arising from the effect of latent variables. Although statistical dependencies do not necessarily imply the existence of causal relationships, partial correlations, which we use here, may provide indications of causal pathways (Epskamp & Fried, 2018). Calculating

directed graphs, which aim to estimate causal relationships directly is an interesting possibility for future work, but it presents its own challenges (Epskamp & Fried, 2018).

In some scenarios a graph theory analysis may be superior to a latent variable characterization. Boorsboom & Cramen (2013) argue that it affords a better understanding of psychopathological conditions. In their view there may not be genuine underlying variables such as depression and anxiety. A more accurate description may be obtained by studying how symptoms interact with each other. For example, insomnia may contribute to fatigue.

Few works have applied graph theory to neuropsychological data. Kellermann et al., (2016) found that networks from patients with temporal epilepsy divide into less communities than those of neurologically healthy controls. Along similar lines, a second population of epilepsy patients, in this case pediatric, was found to have lower functional integration and segregation than healthy controls (Garcia-Ramos et al., 2015). These studies illustrate how network analysis can characterize the balance between the segregation and integration of cognitive domains.

Other advantages of graph theoretic tools include the following (Rubinov & Sporns, 2010). A two-dimensional graph, or layout, can be calculated where nodes (neuropsychological test scores in the present study) are placed so that the distance between them reflects their statistical relationship, providing a rich description of dependencies between variables. Second, a community analysis can divide nodes into categories for which intragroup relationships are strongest and intergroup dependencies are weakest. Third, graph theory allows calculating indices of network organization, such as node centrality, which may help to identify the most important nodes, in the sense that they are more associated with

other nodes, and their disruption may have the strongest influence in on the whole network, as previously demonstrated in domains as diverse as airports and neural networks (Bing, 2014; Borsboom & Cramer, 2013; Rubinov & Sporns, 2010). Bing (2014) showed that removing a few central nodes in an airport network can severely affect the overall connectivity. In psychopathology, a study suggested that central symptoms are better able to predict the onset of depression than peripheral ones (Boschloo et al., 2016). In addition, simulations of networks of psychopathological symptoms (Robinaugh et al., 2016) show that intervention in central nodes have the strongest influence on the rest of the network, especially when the graphs are composed exclusively of positive edges, as is our case. All these results suggest that central nodes could be most influential also in neuropsychological networks, but this should be investigated in future work.

Some centrality measures, such as strength, are sensitive to local connections while others such as closeness and betweenness, indicate how the node is connected globally. Finally, another measure of interest is small-worldness, which quantifies to what extent networks show both high local clustering and high global efficiency, a combination which might reflect a simultaneous capacity for integration and segregation of information.

In the present study we aimed to apply, for the first time, graph theory to neuropsychological data from healthy, cognitively impaired older adults and Alzheimer's Disease patients. We took advantage from the publicly available National Alzheimer's Coordinating Center (NACC) database, which gathers neuropsychological data from participants who range from cognitively normal to mildly cognitively impaired to demented, collected by means of a standardized clinical evaluation. (Beekly et al., 2007; Morris et al., 2006; Weintraub et al., 2009). A previous attempt to characterized the statistical interdependencies among neuropsychological scores from this database (Hayden et al., 2011) revealed that the factorial structure obtained with a latent variable analysis coincides with the

organization proposed beforehand based on theoretical grounds (Weintraub et al., 2009). Specifically, the identified factors correspond to attention, processing speed/executive function, language, and memory (Table 1.). The main result of this previous study is that this same structure is preserved for the three diagnostic groups, namely, cognitively healthy controls, patients with mild cognitive impairment and patients with dementia.

On the other hand, extensive research in developmental psychology and aging indicates that cognitive ability gradually segregates into distinct aptitudes from childhood to adulthood (differentiation), and then becomes again dedifferentiated as we age (de Frias, Lövdén, Lindenberger, & Nilsson, 2007; Gonzalez-Burgos, Hernández-Cabrera, Westman, Barroso, & Ferreira, 2019; Li et al., 2004; Wilson, Segawa, Hizek, Boyle, & Bennett, 2012).

The present analysis may therefore help to clarify if the network structure is similar in cognitively healthy older adults and dementia patients as suggested by Hayden et al., (2011) or cognitive functions became more dedifferentiated in the patient population.

The main goals of the present study are: 1) to characterize the network structure of the neuropsychological functions as reflected by test scores in three populations of older adults with healthy cognition, cognitive impairment, and dementia. 2) to provide a ranking of cognitive functions according to centrality. 3) to serve as a proof of concept for the application of network analysis to neuropsychological data. No hypothesis was formulated with respect to whether there would be between-group differences in neuropsychological network structure. With regard to the centrality of neuropsychological functions, we hypothesized that executive functions would be at the top of the ranking, reflecting the role they are typically assigned in controlling and regulating other cognitive processes (Baddeley, 1996). No additional hypotheses were formulated with respect to the other cognitive functions. Finding whether the network organization and cognitive ranking is common or

different across diagnostic groups in older adults could have implications for diagnosis and treatment.

## **Methods**

An important consideration when characterizing data using network analysis is how accurate and stable the results are. We employed recently developed tools to specifically provide measures of network accuracy (i.e., how prone they are to sampling variation), stability (with respect to the number of samples) and significance of the differences across network connections and centrality estimates (Epskamp et al., 2018). Although these methods have been developed recently they have already extensively been used in other fields (Bryant et al., 2017; Levinson et al., 2017; Mitchell et al., 2017). We additionally assessed the replicability of the results with respect to data sampling, sample sizes and parameter choice.

### ***Neuropsychological Tests***

The Uniform Data Set Neuropsychological Battery (Weintraub et al., 2009) employed in the present study includes 8 neuropsychological tests and 12 scores (Table 1.): Logical Memory Story A (Immediate and Delayed recall) from the Revised Wechsler Memory Scale (Wechsler, 1987), Boston Naming Test (Goodglass H., Kaplan E., 2001), Digit Symbol from the Revised Wechsler Adult Intelligence Scale (Reitan & Heinrichs, 1993), Trail Making Test Parts A and B (Reitan & Heinrichs, 1993), Digits Forward and Digits Backward from the Revised Wechsler Memory Scale (Wechsler, 1973) and semantic fluency (Animals and Vegetables (Morris et al., 1989)). The present analysis was approved by the review board of the School of Psychology of the Complutense University of Madrid. Differences in neuropsychological scores and demographic variables between excluded and included

participants and between diagnostic groups were measured with two-sample t-tests for all variables except for variables race and sex for which Fisher's exact test was employed.

[INSERT TABLE 1 HERE]

### ***Grouping of subjects***

As in Hayden et al. (2011), the global score from the Clinical Dementia Rating (Morris, 1993) was used to classify participants into groups with CRD=0 as healthy controls, CDR=0.5 as Mild Cognitive Impairment patients and CRD>0.5 as Dementia Patients. The Clinical Dementia Rating assesses 6 cognitive domains including memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care.

### ***Inclusion criteria***

Participants were 55 or older at the time of the first visit, had completed at least two visits, reported English as their first language and their 12 neuropsychological measures were available. The first study visit was used in the analyses. Cases were not required to have the same CDR score at the two visits they completed. Each NACC Center has its own protocol to recruit participants. The samples are best described as referral-based or volunteer case series.

### ***Preprocessing***

As in Hayden et al. (2011), data were normalized using the Blom transformation (Ludwig, 1961) because some of the test scores showed ceiling and floor effects. Scores for the Trail Making Test, Parts A and B reflect the time employed to complete the task. These scores were inverted, multiplied by -1, so that higher scores reflect better performance for all tests. The same approach was employed by Kellermann et al. (2016).

### *Measuring correlations*

One of the most common approaches to quantify the statistical relationships between variables is Pearson's correlation coefficient. A limitation is that spurious correlations are introduced, for instance, when two variables are not directly related to each other, but are connected through a third one. Partial correlations mitigate this problem by calculating the statistical dependencies between two nodes while controlling for the influence of the rest of the nodes (Epskamp & Fried, 2018). An added benefit of partial correlations is that they reduce the task impurity problem. The disadvantage of partial over full correlations is that estimations can be inaccurate because the number of variables is typically large in relation to the available empirical data. This second problem is addressed with the use of regularization, which favours sparse solutions, that is, solutions where a large fraction of the partial correlations between pairs of variables is zero. Partial correlations with regularization were calculated with the function `estimateNetwork` from R (R Core Team, 2018) package `bootnet` (Epskamp et al., 2018), with option "EBICglasso". `bootnet` uses R package `ggplot2` (Wickham, 2016) for data visualization. Function `estimateNetwork` selects the optimal amount of regularization according to the extended Bayesian information criterion. In summary, partial correlations with regularization measure how statistically related are pairs of score tests while controlling for the effect of other test scores. They are the basis for obtaining a network that reflects the correlational organization of the set of tests. The nodes of the networks are the test scores. The connections between nodes are referred to as links and the corresponding value of the partial correlation is termed weight.

### *Accuracy and stability of the analysis*

Data from each diagnostic group was randomly divided into 2 subgroups to produce two separate subfigures and assess the replicability of the results with respect to the data sample used. The two subgroups within each group were labelled A and B. We also explored how robust the results were with respect to imposing a threshold in the values of the partial correlations. Whether a threshold is used and how it is set can have a large effect on the resulting network and its properties. As a result of using regularization in the calculation of partial correlations, a fraction of links already had a weight value of 0. This was the reference result. We investigated the effect of imposing an additional threshold. Another possible source of variability in the results is the number of subjects available for each group, that is, the sample size. In particular, between-group differences in sample sizes may introduce spurious between-group differences in network structure. To control for this, we obtained additional results where the sample sizes had been matched. A fourth parameter, the repulsion force (Fruchterman & Reingold, 1991) used to obtain the network graphs in Figure 1, was also systematically investigated. The Fruchterman & Reingold algorithm imposes an attraction force between nearby nodes and a repulsion force between all nodes to calculate their position in the graph.

### ***Network graphs***

The network graphs (Figure 1) were calculated with the R function EstimateNetwork (Epskamp et al., 2018) employing the option “layout=spring”. The algorithm finds spatial locations so that strongly correlated nodes are placed together (Fruchterman & Reingold, 1991).

### ***Within-group statistical contrasts between link weights as measure of network accuracy***

One of the measure of the accuracy of a network characterization proposed in Epskamp et al. (2018) is the within-groups statistical contrast of link weights (partial correlations in our case), which indicates how likely it is that the stronger links will remain so for a different sampling of the same population of subjects. This analysis is provided in the Appendix.

### ***Between-group statistical contrast of link weights.***

To assess how different the networks corresponding to different diagnostic groups were, between-group statistical tests of link weights were calculated using an independent samples t-test with permutations plus correction for multiple comparisons. The null distribution was calculated with permutation tests to avoid assumptions about data normality. Five thousand permutations were employed. Multiple comparison correction was carried out with false the discovery rate (Benjamini & Hochberg, 1995), with  $q=0.05$ . Distributions to allow the calculation of the statistical tests were obtained by dividing the data from each diagnostic group into 10 subgroups, and correlation matrices were calculated for each subgroup. The corresponding graphs were arranged in a circle to facilitate the visual comparison across groups. An additional advantage of permutation testing is that it avoids the false detection of differences between groups arising from differences in sample sizes by constructing surrogate datasets which replicate the number of sample units in the original datasets. These potential differences would then also be present in the null distribution.

### ***Communities***

Classification of nodes into communities was performed using function *community\_louvain* from the brain connectivity toolbox (Rubinov & Sporns, 2010). The algorithm attempts to find a classification that maximizes the weight of the intra-community links while minimizing the weight of the inter-community links. As in the previous section, data for each

diagnostic group was divided into 10 subsamples. A graph for communities was created (Figure 3) where the link width indicates the fraction of times, out of 10, that a given pair of nodes is classified into the same community. The chosen graph layout was circular.

### *Graph-theoretic indices*

Graph-theory provides several indices that can help to determine the relative relevance of the different nodes. We calculated three commonly used centrality measures: betweenness, closeness and strength (Rubinov & Sporns, 2010). Function `centralityPlot` from the R package `qgraph` (Epskamp et al., 2018) was employed to calculate these measures (Figure 4).

Centrality indices quantify how connected a node is to other nodes. In our case this reflects how strong the statistical dependencies between a given node and the rest of the nodes are. Strength centrality quantifies how strong the dependencies are with nodes with which there is a direct edge. The other two measures include the effect of dependencies when more than one step is considered, and therefore include indirect connections. The ‘distance’ between two nodes decreases as the number of edges needed to go from one node to the other decreases, and as the weight of the connecting edges increases. Closeness centrality measures the mean shortest distance from a given node to the rest of the nodes. The closer a node is to the rest of the network the more likely it is that a change in this node would influence the network globally. As for betweenness centrality, it measures the fraction of times that a node is in the shortest path between pairs of other nodes. It has been proposed that in psychopathology networks it could reflect to what extent the node plays a transmission/mediating role (Frewen et al., 2013) between symptoms (e.g. worrying indirectly causing fatigue, via sleeping problems). In neuropsychology, if a given test has high betweenness centrality, it would be more likely to play a mediating role between pairs of other tests.

To assess the stability of the results, the dataset from each diagnostic group was randomly divided into two subsets, A and B. A second estimation of the robustness of the centrality indices was derived as proposed in Epskamp et al. (2018). This second method calculates a distribution of indices by bootstrapping, that is, by randomly choosing subsets of subjects. It is described in the Appendix.

### ***Small-worldness***

Networks with large small-worldness index might indicate both high values of functional integration and segregation (Rubinov & Sporns, 2010). The index was calculated following Humphreys and Gurney (2008). First, the link weights which were different from zero after regularization were binarized to 1. Then, the small-worldness index was obtained as the product of the relative clustering and the relative efficiency, where both terms were normalized with respect to a distribution of values obtained from randomly connected networks with the same number of nodes and links. For realizations for which the graph was not fully connected the largest connected component was employed. Statistical significance was calculated with a two-samples t-test between all pairs of diagnostic groups for the 3 measures separately (small-worldness, relative-clustering, and relative efficiency) where the null-distributions were calculated with permutation testing, and the 9 statistical tests were subsequently jointly corrected for multiple comparisons with the false discovery rate with  $q=0.05$ .

### ***Centrality ranking of cognitive functions***

To rank the cognitive functions according to centrality, a value for each measure (closeness and betweenness) and cognitive function was obtained by averaging across diagnostic groups

(control, mci and dementia), subgroups (A and B) and neuropsychological test scores. Then the average values for closeness and centrality were combined with the geometric mean as they have different ranges. As the ranking is obtained by aggregating the centrality measures from the test scores, and these are subject to a certain degree of task impurity, the resulting measure will also be affected. We comment on this in the discussion section.

## **Results**

### ***Demographics and neuropsychological scores***

A total of 19317 participants over the age of 54 and reporting English as their primary language had completed the neuropsychological assessment (form C1) at least during two visits from September 2005 to August 2017. The dataset included data from 35 Alzheimer's Disease Centers. After excluding those participants for which not all the 12 neuropsychological measures were available the number was reduced to 15644. The groups classified according to the Clinical Dementia Rating, were cognitively healthy older adults, CDR=0, n=7623; Mild Cognitive Impairment patients, CDR=0.5, n=5981; and dementia patients, CDR $\geq$ 1, n=2040. 83% of dementia patients had a diagnosis of Alzheimer's Disease. Table 2 provides the mean and standard deviation of the demographic variables for the different groups, while Table 3 contains the descriptive statistics of the neuropsychological scores. When comparing the excluded group with the group fulfilling the inclusion criteria the statistical contrast of demographic variables in Table 2 and neuropsychological scores in Table 3 yielded significant differences ( $p < 0.001$ ) for all variables except for sex, education and race (unpaired t-test). When comparing the 3 different diagnostic groups (control participants, MCI patients and dementia patients) all statistical contrasts with respect to the demographic variables (Table 2) were significant ( $p < 0.001$ ), except for race and education

when comparing the MCI and dementia groups, which were not significant, and sex when comparing the MCI and dementia groups which was significant with  $p < 0.01$  (unpaired t-test). All statistical contrasts between diagnostic groups with respect to the neuropsychological scores were significant.

[INSERT TABLES 2 AND 3 HERE]

### *Network graphs*

After calculating the partial correlations with regularization, the proportion of links with absolute values different from zero was 45%, considering the three diagnostic groups. The results reported in the different figures correspond to networks for which no additional threshold was imposed. In addition, we investigated networks for which an additional threshold was set so that the corresponding fraction of links with absolute weight value above zero was 35% and 25%, as compared to 45% when only regularization is used.

Network graphs for the three diagnostic groups are shown in Figure 1. Within each diagnostic group, A and B denote two randomly drawn data subsamples used to assess the robustness of the results. Colors indicate factor as in Hayden et al. (2011). The first result of interest is that the layouts preserved a similar structure across diagnostic groups. Second, the most central functions were executive functions and language (blue and yellow nodes, respectively). Attention (red) was placed close to executive functions and memory (green) next to language. These two results were robust with respect to sampling as demonstrated by comparing subsamples A and B. Moreover, both outcomes also held when introducing a threshold for the links weights and when using samples of the same size for the different diagnostic groups. Furthermore, the results were stable with respect to the attraction/repulsion balance used in the Fruchterman and Reingold algorithm if the repulsion parameter was larger than 1.2. Values of 1.5, 2.0 and 2.5 were systematically explored. A

value of 1.5 was employed to obtain Figure 1. Finally, the graphs were also robust with respect to where the nodes were initially positioned before applying the algorithm. Randomizing the initial positions did not alter the main conclusions drawn from the layouts. We should note that the layouts provide a visualization of the network and are useful in assisting interpretation. Nevertheless, the main results concerning the centrality of nodes and functions are provided by the quantitative analysis displayed in Figures 6 and 7.

[INSERT FIGURE 1 HERE]

An analysis of the accuracy and stability of the links is provided in the Appendix. Figures A1 and A2, in the Appendix, indicate that clear within-group weight differences exist between sets of links.

### ***Between and within factor links***

Figure 2 shows the distribution of link weights from the networks depicted in Figure 1 divided into those connecting neuropsychological tests belonging to the same (within) or different (between) factors. As expected, within-factor connections were stronger than between-factor connections. Nevertheless, between-factor connections represented a sizable contribution to the global distribution. Similar results were found when changing the threshold value of the partial correlations or employing the same sample size for the three diagnostic groups.

[INSERT FIGURE 2 HERE]

### ***Differences between diagnostic groups***

A between-group statistical comparison of link weights (data not shown) suggests that there are stronger connections between “temporal lobe functions” (memory and language) and

attentional-executive functions in the dementia than in the control group, with the MCI group occupying an intermediate position. Nevertheless, the replicability of the results with respect to threshold changes and sample size matching was only moderate.

### *Community analysis*

Figure 3 displays a community analysis for the three diagnostic groups. The width of the links denotes how often a given pair of nodes was classified as belonging to the same community. Nodes were arranged in a circle. Results for the 3 groups were generally consistent with factor analysis (Hayden, 2011), in that links are stronger within factors, with some small differences. First, the attention tests appeared divided into two communities. This may reflect the fact that the backward span tasks impose additional demands engaging working memory rather than short term memory. Second, in a small fraction of instances, language and executive functions in control participants, and language and memory in dementia patients were classified into the same community. These results were stable with respect to threshold choice and sampling size matching.

[INSERT FIGURE 3 HERE]

### *Centrality measures*

Figure 4 shows results for 3 graph-theoretic indices (betweenness, strength and closeness) for the different neuropsychological tests and diagnostic groups. The upper subfigures correspond to the A subgroups, with the bottom panels representing the B subgroups. Vertical lines delineate the border between functions. The measures characterizing how central a node is are betweenness and closeness. Combining these two types of information it is apparent that the most central function was executive function followed by language. This agrees with the results obtained with the network graphs (Figure 1). At the node level, the most central node was the Trail Making Test Part B (TMT-B). The strength index, which quantifies direct

connections, did not show large variations across nodes, except for the fact that the language nodes presented lower values. These results were robust with respect to data sampling as shown in the Figure and when changing threshold or matching sample sizes. Additionally, Figure A1 in the Appendix is part of the accuracy and stability analysis proposed by Epskamp et al., (2018) and indicate that the centrality results were robust.

[INSERT FIGURE 4 HERE]

The values reported in Figure 4 have been averaged according to function in Figure 5 to provide a summarized view. Closeness provides the most robust measure of centrality with executive function as the most central function followed by language. Betweenness yields a slightly different picture, with executive functions still at the top followed by attention. These results were robust with respect to threshold change and sample size matching.

[INSERT FIGURE 5 HERE]

### ***Small-worldness***

The three diagnostic groups had similar values for the three measures (data not shown). From the nine statistical contrasts performed (three group pairings times three measures) some statistically significant differences between relative clustering measures were found, but this was not stable with respect to changes in threshold or sample size.

### ***Centrality ranking of cognitive functions***

The average centrality values for each cognitive function across diagnostic groups and nodes are 0.275 for executive functions, 0.174 for attention, 0.165 for language and 0.128 for memory.

## **Discussion**

We investigated with network analysis the structure of the relationships between neuropsychological test scores in populations of healthy older adults and Mild Cognitive Impairment and dementia patients. One of the goals of the study was to serve as a proof of principle for the application of graph theory to neuropsychological data. The use of network analysis allows exploiting tools such as centrality indices and graphs of dependencies, but it also presents challenges for the interpretation of results such as the effects of task impurity, the dependency of the networks on the chosen or available tests and the usefulness of the derived indices in clinical practice. These issues are discussed below.

### *Central functions and tests*

Central nodes could be more important than peripheral ones as they are more connected directly or indirectly to other nodes and their degradation is likely to have a larger effect on the overall network than peripheral nodes. According to the centrality measures closeness and betweenness the order in cognitive functions from most to least central is executive function, followed by attention, language, and memory. This illustrates how network analysis can yield a quantitative hierarchy of cognitive functions. Although executive functions are commonly assumed to be at the top of the hierarchy, this has not been quantified before.

This quantitative analysis is complemented by the qualitative graphs in Figure 1. Memory and attention are found in the periphery, with memory linked to language and attention to executive function. Therefore, attention and language swap positions in the graphical representation. It should be noted, nevertheless that the values for memory and attention are similar, and if we consider closeness as centrality measure, which is more reliable than betweenness, rather than the average of closeness and betweenness, language recovers the second position. It follows that language and attention probably deserve a similar centrality status in our analysis.

Figure 4 shows the centrality measures at the single score level. The picture for closeness remains like that of Figure 5 where measures were averaged according to function, as scores belonging to the same function behave similarly. Therefore, the centrality ranking with executive function at the top, followed by language, attention and memory is preserved. Betweenness shows more within-function variability. Combining the two types of information the most central test is the Trail Making Test – Part B. Graph theoretic indices that assess centrality may help to identify the most relevant neuropsychological tests, both in the context of understanding the cognitive system at a fundamental level, and in terms of guiding neuropsychological assessment and intervention.

The present analysis therefore suggests that executive functions play a central role both in healthy ageing and dementia. Measures of executive function correlate with scales of instrumental activities of daily living in older adults at risk for cognitive decline (Jefferson et al., 2006) and predict longitudinal changes across the spectrum from healthy aging, through Mild Cognitive Impairment to dementia (Tomaszewski Farias et al., 2009) suggesting a key role in the organization of adaptive behavior across health and disease. The Trail Making Test - Part B has the highest centrality values. The general view is that this test measures cognitive flexibility (Arbuthnott & Frank, 2000; Lamberty & Adams, 1994; Pontius & Yudowitz, 1980), but see Kortte, Horner and Windham (2002) for an alternative perspective. Flexibility plays a fundamental role in cognition as we need to adapt to a continuously changing environment. In addition to cognitive flexibility, the TMT-B test may also capture processing speed, which may also play an important role in cognition for older adults (Salthouse, 1996). Language occupies the third position in the centrality ranking, and its nodes are placed in the proximity of the executive function nodes. This may result from the fact that language here is represented by verbal fluency and lexical access and these two processes rely on executive function and processing speed. Several previous results support

the notion that lexical access and fluency play key roles in cognitive function. Fluency tests reflect how we organize our thinking to produce clusters of related words (Estes, 2006).

Moreover, reduced category fluency has been observed in patients with Alzheimer's Disease (Rascovsky et al., 2007). Likewise, lexical access and verbal fluency deficits are common in Alzheimer's Disease and Mild Cognitive Impairment (Taler & Phillips, 2008).

Three different elements were employed to ensure that the results we obtained are robust: 1) We used the approach developed by Epskamp et al., (2018) specifically to assess the accuracy of the results of network analysis. 2) We divided the dataset into two subsets. 3) We explored the space of parameter values.

An important question for future work is to what extent the ranking found in the present work changes when including additional neuropsychological test to the analysis, covering other aspects of cognition. Presently we have considered twelve neuropsychological measures grouped into four functions. Future work could include tests measuring inhibition, perceptual processes, motor control and visuospatial skills. In general, an analysis at the observable level will be more dependent on the available neuropsychological tests, even if they represent the same cognitive functions as in the present analysis, than the equivalent analysis at the latent level, which should be more invariant with respect to the particular dataset employed. An important question for future research is how stable key results such as centrality rankings are when considering different sets of tests.

Another natural extension would be to characterize sets of neuropsychological test scores from younger populations. For example, latent variable analysis has uncovered that the rate of development of different executive function components varies (Huizinga et al., 2006). Along the same lines, the different components of executive functions are closely related until 9 years of age and become more separable after that (Brydges et al., 2014). It would be interesting to investigate if a network analysis provides a compatible picture.

A similar question is to what extent the present centrality hierarchy established according to centrality measures is compatible with an ordering of cognitive functions obtained with other measures. Here we employed the network measures closeness and betweenness centrality. Other network centrality measures are within-module degree z-score and the participation coefficient (Rubinov & Sporns, 2010). Outside network analysis an approach which offers some kind of hierarchy of variables is nested factor analysis (Gignac, 2005). An important difference nevertheless is that the nesting occurs for the latent, as opposed to observable, variables.

Criteria to choose which test to use in a clinical setting when the assessment must be done in a short time period may include the number of cognitive processes it measures or how well it reflects everyday cognitive demands. For example, the trail making test is sometimes used to assess driving fitness (Dobbs & Shergill, 2013). Centrality is likely to be related to these two criteria. A test which recruits many cognitive functions would probably have a high strength value as this measure reflects direct connections. This will also have an impact on closeness and betweenness which are also affected by direct connections, but the correlation would be less strong, as these tests are also sensitive to indirect connections. Importantly though, network analysis provides a quantitative and well-defined stipulation of how to calculate centrality. While how many aspects a test reflects or how well it characterizes daily activities are relevant notions, they need to be operationalized to be used in research and clinical practice. Future work could compare centrality as a predictor of clinical relevance with other possible criteria such as the correlation between individual tests and the mean of the group of tests reflecting a given cognitive function, global measures such as the MMSE score, or the extent to which the individual test loads onto multiple cognitive factors.

Previous results (Bing, 2014; Boschloo et al., 2016; Robinaugh et al., 2016) suggest that central nodes are more influential than peripheral nodes. Future work could whether this is also the case in neuropsychology. For example, a walking version of the trail making test has been proposed as a possible tool for the early detection of cognitive impairment (Perrochon & Kemoun, 2014). The clinical relevance of centrality indices could be assessed by calculating if central tests predict better the onset of dementia than peripheral ones as has been found for psychopathology symptom and depression onset (Boschloo et al., 2016)

We defined the centrality of a cognitive function as the aggregate centrality of its composing neuropsychological test scores. As the individual tests are subject to some degree of task impurity, this will also be reflected in the global index. Future work could investigate combining the advantages of graph theory and factor analysis by, for example, weighting the centrality indices of the tests with the loadings obtained from factor analysis. It should be noted though that some degree of impurity is to be expected in latent variable analysis also, because of the approximations involved, and the fact that there are genuine dependencies between cognitive functions.

### ***Changes in network structure as the disease progresses***

An important issue is whether the interrelationship between neuropsychological functions changes with the course of the condition. Previous results from factor analysis (Hayden et al., 2011) suggested that factors were stable across diagnostic groups and that the factorial structure agreed with the predefined theoretical classification. The present analysis is in general consistent with this view. First, the distances between nodes belonging to the same factor are shorter than those between nodes belonging to different cognitive functions (Figure 1). Also, the community analysis in Figure 3 is generally consistent with the factorial characterization. Second, the structure of the network graphs (Figure 1), the graph theoretic

measures (Figure 5) and small-world indices are largely stable across diagnostic groups. Small between-group differences are found with respect to the division into communities (Figure 3) and link weights but stability as dementia sets in dominates. On the other hand, Alzheimer's Disease has been conceptualized as a disconnection syndrome, where a progressive loss of synaptic connections between brain regions impairs long distance communication leading to cognitive decline (Selkoe et al., 2012). Along this line, neuroimaging studies have shown an association between reductions in both anatomical and functional connectivity and cognitive impairment (Garcés et al., 2014). Future work could combine the characterization of neuropsychological and neuroimaging data from a network perspective.

In contrast to the present results, patients with temporal epilepsy appear to have a more compact community organization than healthy controls (Kellermann et al., 2016). Nevertheless, Alzheimer's Disease and epilepsy are etiologically different, with long distance connections most affected in the former. Likewise, the current network analysis did not identify evidence of dedifferentiation of cognitive functions with the onset of mild cognitive impairment and dementia. Although dedifferentiation has been described with aging (de Frias et al., 2007; Wilson et al., 2012) rather than in dementia patients, between-group differences could have been expected if dementia resembles accelerated aging (Dras & Blumenthal, 1992).

***Does graph theory provide complementary information to factor analysis on the organization of cognitive processes in older adults?***

One important methodological question is what the graph theory approach adds to the characterization of the statistical structure of the data offered by a latent variable analysis. As we discussed in the previous sections, the results we obtained with the network graphs and

graph-theoretic indices offer a complementary description to the latent variable view in that they can characterize which functions and tests are most central.

From a more fundamental point of view, we may ask whether there are genuine statistical dependencies at the level of observable variables not explained by the influence of the latent factors. A structure dominated by factors is likely to yield stronger intra than inter-factor links and would probably display low variability across intra-factor links (Epskamp, Kruis, & Marsman, 2017). The fact that in the present study there is variability in link strength and inter-factor links have values in the same range as intra-factor links (Figure 2) suggests that the statistical dependencies at the observable level constitute a relevant contribution to the network structure.

In addition, from a theoretical point of view, a latent variable model may be more appropriate than network analysis when there are underlying causes of the symptoms (Borsboom & Cramer, 2013) while the latter may be preferable when that is not the case. These authors argue that in the case of medical conditions, such as cancer or the flu, the more useful object of study is the etiological cause. In contrast, they maintain that such a model has not proved fruitful enough for psychopathology as there appears to be no clear underlying cause for conditions such as major depression and anxiety. In this case, they hold that a network analysis perspective is better suited. Although both views can be complementary in most situations, an important question is whether the organization of cognitive systems, as measured by neuropsychological tests, has a closer resemblance to the medical model or to the structure Borsboom and Cramer advocate for psychopathological conditions. Executive functions show elements supporting both interpretations, as reviewed by Friedman & Miyake (2017). In favour of the latent perspective, executive functions are separable when measured with latent variables, are highly heritable at the latent level and activate specific neural areas. On the other hand, in favor of the network analysis view, executive functions are robustly

correlated, appear highly polygenic and activate common neural areas. Taking together these considerations and the results of the present study, it is unlikely that in neuropsychology all interactions occur at the test level, with no contribution from underlying variables, as proposed by authors such as Borsboom & Cramer (2013) for psychophysiology. We nevertheless think that these types of dependencies constitute a relevant element of the cognitive structure, and that network analysis provides a fruitful view for neuropsychology.

From a complementary perspective, descriptions in terms of both observable and latent variables provide a very approximate and simplified description of a system, particularly in cases where the object of interest, such as the brain, is very complex. A latent variable approach, if successful, has the advantage of providing a more compact and invariant representation. This may include the description of the dependencies between the different parts of the system. In our case these could be the different cognitive functions. But working at the observable level has its own strengths. As the data is less processed, the information loss is reduced. In addition, the larger number of variables makes it possible to use the powerful tools developed within graph theory, including centrality indices and the plotting of graphs. For network analysis to really be useful a minimum number of variables is required. Both descriptions, network and latent variable analysis, can therefore complement each other. Future work could explore carrying out a network analysis at the latent variable level if the dataset allows for the identification of a large enough number of such variables.

### ***Limitations and future work***

The group of patients diagnosed with dementia was composed mainly by Alzheimer patients (83%), but it also included cases with other types of dementia. Along the same lines, some

patients with a CDR score of 0.5 may have a different etiology than suspected neurodegenerative disease. Future work could analyze the different subgroups separately.

Two limitations of the NACC database are that the sample is clinical, and therefore some differences will exist with respect to the general population. Also, many sites contribute to the database, which is likely to introduce some heterogeneity despite the effort to coordinate tests and procedures. On the other hand, it is precisely the large number of contributing sites that allow for the large sample size which is a strong asset for this type of analysis. Along similar lines, as shown in Table 2, some demographic differences exist across diagnostic groups. The variable with sizeable differences is sex. Again, the choice was to focus on having large samples. Here we provided a cross-sectional analysis. Future work could investigate the trajectory of the participants in this database as they age.

Neuropsychological tests may measure different levels of cognitive functioning, with some tests tapping into more general domains than others. An interesting question is how sensitive is network analysis to the level of homogeneity of the nodes. This is an issue which is also relevant in psychopathology and personality research, where network analysis has been widely used. If a large number of tests are available, one could choose tests at similar levels of description. On the other hand, comparing analyses with homogeneous and heterogeneous levels of descriptions could be informative in itself. Let us take for example an analysis which includes a global test score and the corresponding subscales that compose it. One possibility is that the global score and the subscales collapse to very similar locations in the network graph, which would indicate that some of them are redundant. Another possibility is that some of them take a central role while others appear in the periphery, suggesting that they are less important. In summary, this type of exploration could provide further insights into the structure of cognitive functions. A complementary future avenue

would be to use techniques that aim to combine overlapping nodes when appropriate, as has been proposed for networks based on VAR(1) models (Bulteel et al., 2018)

In the present analysis we calculated three of the most common centrality measures: betweenness, closeness, and strength. Future work could consider additional centrality measures (Rubinov & Sporns, 2010), including those related to modular analysis, especially when a large number of neuropsychological tests are available. Another possible extension is to use directed networks that explicitly aim to estimate causality relationships between nodes. Nevertheless, partial correlations may provide indications of causal pathways and working with directed graphs can be challenging (Epskamp & Fried, 2018).

### ***Conclusion***

In conclusion, characterizing neuropsychological processes from a network perspective can help to understand cognition by revealing which tests and function are most central. The present analysis confirms the fundamental role of executive function in the cognitive organization of older adults and yielded a centrality ranking for the other analyzed functions. This type of approach may therefore assist to refine models of cognitive function and guide cognitive intervention.

### **Declaration of interest**

The authors report no conflict of interest.

The NACC database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIA funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI

Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI M. Marsel Mesulam, MD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG005131 (PI James Brewer, MD, PhD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG053760 (PI Henry Paulson, MD, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), P50 AG047270 (PI Stephen Strittmatter, MD, PhD).

## References

- Arbuthnott, K., & Frank, J. (2000). Trail Making Test, Part B as a Measure of Executive Control: Validation Using a Set-Switching Paradigm. *Journal of Clinical and Experimental Neuropsychology*, 22(4), 518–528. [https://doi.org/10.1076/1380-3395\(200008\)22:4;1-0;FT518](https://doi.org/10.1076/1380-3395(200008)22:4;1-0;FT518)
- Baddeley, A. (1996). Exploring the Central Executive. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 49(1), 5–28.

<https://doi.org/10.1080/713755608>

- Beekly, D. L., Ramos, E. M., Lee, W. W., Deitrich, W. D., Jacka, M. E., Wu, J., Hubbard, J. L., Koepsell, T. D., Morris, J. C., Kukull, W. A., Reiman, E. M., Kowall, N., Landreth, G., Shelanski, M., Welsh-Bohmer, K., Levey, A. I., Potter, H., Ghetti, B., Price, D., ... Raskind, M. (2007). The National Alzheimer's Coordinating Center (NACC) database: The uniform data set. In *Alzheimer Disease and Associated Disorders* (Vol. 21, Issue 3, pp. 249–258). <https://doi.org/10.1097/WAD.0b013e318142774e>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A Practical and powerful approach to multiple testing. *J.Roy.Statist.Soc.*, *57*, 289–300.  
<http://inspirehep.net/record/1625829/>
- Bing, D. (2014). Reliability analysis for aviation airline network based on complex network. *Journal of Aerospace Technology and Management*, *6*(2), 193–201.  
<https://doi.org/10.5028/jatm.v6i2.295>
- Borsboom, D., & Cramer, A. O. J. (2013). Network Analysis: An Integrative Approach to the Structure of Psychopathology. In *Annual Review of Clinical Psychology* (Vol. 9, Issue 1). <https://doi.org/10.1146/annurev-clinpsy-050212-185608>
- Boschloo, L., Van Borkulo, C. D., Borsboom, D., & Schoevers, R. A. (2016). A Prospective Study on How Symptoms in a Network Predict the Onset of Depression. In *Psychotherapy and Psychosomatics* (Vol. 85, Issue 3, pp. 183–184).  
<https://doi.org/10.1159/000442001>
- Bryant, R. A., Creamer, M., O'Donnell, M., Forbes, D., McFarlane, A. C., Silove, D., & Hadzi-Pavlovic, D. (2017). Acute and chronic posttraumatic stress symptoms in the emergence of posttraumatic stress disorder a network analysis. *JAMA Psychiatry*, *74*(2), 135–142. <https://doi.org/10.1001/jamapsychiatry.2016.3470>
- Brydges, C. R., Fox, A. M., Reid, C. L., & Anderson, M. (2014). The differentiation of

executive functions in middle and late childhood: A longitudinal latent-variable analysis.

*Intelligence*, 47, 34–43. <https://doi.org/10.1016/j.intell.2014.08.010>

Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2018). Improved Insight into and Prediction of Network Dynamics by Combining VAR and Dimension Reduction. *Multivariate Behavioral Research*, 53(6), 853–875.

<https://doi.org/10.1080/00273171.2018.1516540>

Contreras, A., Nieto, I., Valiente, C., Espinosa, R., & Vazquez, C. (2019). The study of psychopathology from the network analysis perspective: A systematic review. In *Psychotherapy and Psychosomatics* (Vol. 88, Issue 2, pp. 71–83). S. Karger AG.

<https://doi.org/10.1159/000497425>

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. J. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality*, 54, 13–29.

<https://doi.org/10.1016/j.jrp.2014.07.003>

Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., & Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, 26(4), 414–431. <https://doi.org/10.1002/per.1866>

de Frias, C. M., Lövdén, M., Lindenberger, U., & Nilsson, L. G. (2007). Revisiting the dedifferentiation hypothesis with longitudinal multi-cohort data. *Intelligence*, 35(4), 381–392. <https://doi.org/10.1016/j.intell.2006.07.011>

Dobbs, B. M., & Shergill, S. S. (2013). How effective is the trail making test (parts a and b) in identifying cognitively impaired drivers? *Age and Ageing*, 42(5), 577–581.

<https://doi.org/10.1093/ageing/aft073>

Dowling, N. M., Hermann, B., La Rue, A., & Sager, M. A. (2010). Latent Structure and

Factorial Invariance of a Neuropsychological Test Battery for the Study of Preclinical Alzheimer's Disease. *Neuropsychology*, 24(6), 742–756.

<https://doi.org/10.1037/a0020176>

Dras, D. D. V., & Blumenthal, H. T. (1992). Dementia of the Aged: Disease or Atypical Accelerated Aging? Biopathological and Psychological Perspectives. *Journal of the American Geriatrics Society*, 40(3), 285–294. <https://doi.org/10.1111/j.1532-5415.1992.tb02084.x>

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, 50(1), 195–212.

<https://doi.org/10.3758/s13428-017-0862-1>

Epskamp, S., & Fried, E. I. (2018, July 5). A Tutorial on Regularized Partial Correlation Networks. *Psychological Methods*. <https://doi.org/10.1037/met0000167>

Epskamp, S., Kruis, J., & Marsman, M. (2017). Estimating psychopathological networks: Be careful what you wish for. *PLOS ONE*, 12(6), e0179891.

<https://doi.org/10.1371/journal.pone.0179891>

Estes, W. K. (2006). Learning theory and intelligence. *American Psychologist*, 29(10), 740–749. <https://doi.org/10.1037/h0037458>

Fournier-Vicente, S., Larigauderie, P., & Gaonac'h, D. (2008). More dissociations and interactions within central executive functioning: A comprehensive latent-variable analysis. *Acta Psychologica*, 129(1), 32–48.

<https://doi.org/10.1016/j.actpsy.2008.04.004>

Frewen, P. A., Schmittmann, V. D., Bringmann, L. F., & Borsboom, D. (2013). Perceived causal relations between anxiety, posttraumatic stress and depression: Extension to moderation, mediation, and network analysis. *European Journal of Psychotraumatology*, 4(SUPPL.). <https://doi.org/10.3402/ejpt.v4i0.20656>

- Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. In *Cortex* (Vol. 86, pp. 186–204). Elsevier. <https://doi.org/10.1016/j.cortex.2016.04.023>
- Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph Drawing by Force-directed Placement. *SOFTWARE—PRACTICE AND EXPERIENCE*, 21(1), 1129–1164. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.13.8444&rep=rep1&type=pdf>
- Garcés, P., Ángel Pineda-Pardo, J., Canuet, L., Aurtenetxe, S., López, M. E., Marcos, A., Yus, M., Llanero-Luque, M., Del-Pozo, F., Sancho, M., & Maestú, F. (2014). The Default Mode Network is functionally and structurally disrupted in amnesic mild cognitive impairment-A bimodal MEG-DTI study. *NeuroImage: Clinical*, 6, 214–221. <https://doi.org/10.1016/j.nicl.2014.09.004>
- Garcia-Ramos, C., Lin, J. J., Prabhakaran, V., & Hermann, B. P. (2015). Developmental Reorganization of the Cognitive Network in Pediatric Epilepsy. *PLOS ONE*, 10(10), e0141186. <https://doi.org/10.1371/journal.pone.0141186>
- Gignac, G. E. (2005). Revisiting the Factor Structure of the WAIS-R Insights Through Nested Factor Modeling. *Journals.Sagepub.Com*, 12(3), 320–329. <https://doi.org/10.1177/1073191105278118>
- Goodglass H., Kaplan E., B. B. (2001). *Boston Diagnostic Aphasia Examination*. (Third Edit). Pro-Ed.
- Hayden, K. M., Jones, R. N., Zimmer, C., Plassman, B. L., Browndyke, J. N., Pieper, C., Warren, L. H., & Welsh-Bohmer, K. A. (2011). Factor structure of the national Alzheimer's coordinating centers dataset neuropsychological battery: An evaluation of invariance between and within groups over time. *Alzheimer Disease and Associated Disorders*, 25(2), 128–137. <https://doi.org/10.1097/WAD.0b013e3181ffa76d>
- Heyman, A., Peterson, B., Fillenbaum, G., & Pieper, C. (1996). The consortium to establish a

- registry for Alzheimer's disease (CERAD). Part XIV: Demographic and clinical predictors of survival in patients with Alzheimer's disease. *Neurology*, *46*(3), 656–660.  
<https://doi.org/10.1212/WNL.46.3.656>
- Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017–2036.  
<https://doi.org/10.1016/j.neuropsychologia.2006.01.010>
- Humphries, M. D., & Gurney, K. (2008). Network “small-world-ness”: A quantitative method for determining canonical network equivalence. *PLoS ONE*, *3*(4), e0002051.  
<https://doi.org/10.1371/journal.pone.0002051>
- Jefferson, A. L., Paul, R. H., Ozonoff, A., & Cohen, R. A. (2006). Evaluating elements of executive functioning as predictors of instrumental activities of daily living (IADLs). *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *21*(4), 311–320. <https://doi.org/10.1016/j.acn.2006.03.007>
- Kellermann, T. S., Bonilha, L., Eskandari, R., Garcia-Ramos, C., Lin, J. J., & Hermann, B. P. (2016). Mapping the neuropsychological profile of temporal lobe epilepsy using cognitive network topology and graph theory. *Epilepsy and Behavior*, *63*, 9–16.  
<https://doi.org/10.1016/j.yebeh.2016.07.030>
- Kortte, K. B., Horner, M. D., & Windham, W. K. (2002). The trail making test, Part B: Cognitive flexibility or ability to maintain set? *Applied Neuropsychology*, *9*(2), 106–109. [https://doi.org/10.1207/S15324826AN0902\\_5](https://doi.org/10.1207/S15324826AN0902_5)
- Lamberty, G. J., & Adams, K. M. (1994). Derived trail making test indices: A preliminary report. *Neuropsychiatry, Neuropsychology and Behavioral Neurology*, *7*(3), 230–234.  
<https://psycnet.apa.org/record/1995-31624-001>
- Levinson, C. A., Zerwas, S., Calebs, B., Forbush, K., Kordey, H., Watson, H., Hofmeier, S.,

- Levine, M., Crosby, R. D., Peat, C., Runfola, C. D., Zimmer, B., Moesner, M., Marcus, M. D., & Bulik, C. M. (2017). The core symptoms of bulimia nervosa, anxiety, and depression: A network analysis. *Journal of Abnormal Psychology, 126*(3), 340–354. <https://doi.org/10.1037/abn0000254>
- Ludwig, O. (1961). Blom, Gunnar: Statistical estimates and transformed beta-variables. Wiley/New York, Almquist und Wiksell/Stockholm 1958; 176 S., Kr. 20,—. *Biometrische Zeitschrift, 3*(4), 285–285. <https://doi.org/10.1002/bimj.19610030410>
- McNally, R. J. (2016). Can network analysis transform psychopathology? In *Behaviour Research and Therapy* (Vol. 86, pp. 95–104). Elsevier Ltd. <https://doi.org/10.1016/j.brat.2016.06.006>
- Mitchell, K. S., Wolf, E. J., Bovin, M. J., Lee, L. O., Green, J. D., Rosen, R. C., Keane, T. M., & Marx, B. P. (2017). Network models of DSM-5 posttraumatic stress disorder: Implications for ICD-11. *Journal of Abnormal Psychology, 126*(3), 355–366. <https://doi.org/10.1037/abn0000252>
- Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language, 21*(2), 169–183. <https://doi.org/10.1055/s-2000-7563>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex ““Frontal Lobe”” Tasks: A Latent Variable Analysis. *Cognitive Psychology, 41*, 49–100. <https://doi.org/10.1006>
- Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology, 43*(11), 2412–2412. <https://doi.org/10.4067/S0718-34292016005000011>
- Morris, John C., Weintraub, S., Chui, H. C., Cummings, J., DeCarli, C., Ferris, S., Foster, N. L., Galasko, D., Graff-Radford, N., Peskind, E. R., Beekly, D., Ramos, E. M., & Kukull,

- W. A. (2006). The Uniform Data Set (UDS): Clinical and cognitive variables and descriptive data from Alzheimer disease centers. *Alzheimer Disease and Associated Disorders*, 20(4), 210–216. <https://doi.org/10.1097/01.wad.0000213865.09806.92>
- Perrochon, A., & Kemoun, G. (2014). The walking trail-making test is an early detection tool for mild cognitive impairment. *Clinical Interventions in Aging*, 9, 111–119. <https://doi.org/10.2147/CIA.S53645>
- Pontius, A. A., & Yudowitz, B. S. (1980). Frontal lobe system dysfunction in some criminal actions as shown in the narratives test. *Journal of Nervous and Mental Disease*, 168(2), 111–117. <https://doi.org/10.1097/00005053-198002000-00008>
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Rascovsky, K., Salmon, D. P., Hansen, L. A., Thal, L. J., & Galasko, D. (2007). Disparate letter and semantic category fluency deficits in autopsy-confirmed frontotemporal dementia and Alzheimer's disease. *Neuropsychology*, 21(1), 20–30. <https://doi.org/10.1037/0894-4105.21.1.20>
- Reitan, R. M., & Heinrichs, R. W. (1993). *The {Halstead}-{Reitan} neuropsychological test battery: {Theory} and clinical interpretation* (2nd ed.). Neuropsychology Press. <https://searchworks.stanford.edu/view/11561397>
- Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *Journal of Abnormal Psychology*, 125(6), 747–757. <https://doi.org/10.1037/abn0000181>
- Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3), 1059–1069. <https://doi.org/10.1016/j.neuroimage.2009.10.003>
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition.

- Psychological Review*, 103(3), 403–428. <http://www.ncbi.nlm.nih.gov/pubmed/8759042>
- Selkoe, D., Mandelkow, E., & Holtzman, D. (2012). Deciphering Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine*, 2(1), a011460–a011460.  
<https://doi.org/10.1101/cshperspect.a011460>
- Sporns, O. (2018). Graph theory methods: Applications in brain networks. *Dialogues in Clinical Neuroscience*, 20(2), 111–120.  
<https://doi.org/10.31887/DCNS.2018.20.2/OSPORNS>
- Taler, V., & Phillips, N. A. (2008). Language performance in Alzheimer’s disease and mild cognitive impairment: A comparative review. In *Journal of Clinical and Experimental Neuropsychology* (Vol. 30, Issue 5, pp. 501–556).  
<https://doi.org/10.1080/13803390701550128>
- Thurstone, L. L. (1947). Multiple-factor analysis; a development and expansion of The Vectors of Mind. In *Multiple-factor analysis; a development and expansion of The Vectors of Mind*. University of Chicago Press.
- Tomaszewski Farias, S., Cahn-Weiner, D. A., Harvey, D. J., Reed, B. R., Mungas, D., Kramer, J. H., & Chui, H. (2009). Longitudinal Changes in Memory and Executive Functioning are Associated with longitudinal change in instrumental activities of daily living in older Adults. *The Clinical Neuropsychologist*, 23(3), 446–461.  
<https://doi.org/10.1080/13854040802360558>
- Wechsler, D. (1973). *Wechsler Memory Scale-Revised Manual*. Psychological Corp.
- Weintraub, S., Salmon, D., Mercaldo, N., Ferris, S., Graff-Radford, N. R., Chui, H., Cummings, J., DeCarli, C., Foster, N. L., Galasko, D., Peskind, E., Dietrich, W., Beekly, D. L., Kukull, W. A., & Morris, J. C. (2009). The Alzheimer’s Disease Centers’ Uniform Data Set (UDS): The neuropsychologic test battery. *Alzheimer Disease and Associated Disorders*, 23(2), 91–101. <https://doi.org/10.1097/WAD.0b013e318191c7dd>

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>

Wilson, R. S., Segawa, E., Hizel, L. P., Boyle, P. A., & Bennett, D. A. (2012). Terminal dedifferentiation of cognitive abilities. *Neurology*, 78(15), 1116–1122. <https://doi.org/10.1212/WNL.0b013e31824f7ff2>

## **Appendix**

### ***Within-group statistical contrasts between link weights as measure of network accuracy***

This is one of the accuracy measures proposed in Epskamp et al. (2018). They calculate how likely it is that the stronger links will remain so for a different sampling from the same population of subjects. We therefore assessed, within each diagnostic group, how different the weights of the various links were, by carrying out statistical tests and calculating the 95% confidence intervals. The distribution to obtain the confidence interval was obtained by drawing subsamples of subjects from the original sample (Epskamp et al., 2018), a process referred to as bootstrapping. The statistical test was carried out by calculating the difference between bootstrapped values of two link weights, which then allow to construct a bootstrapped confidence interval of the difference. The difference is significant if zero is not within this confidence interval. This analysis revealed that a large fraction of link pairs had statistically significant differences in weight (data not shown), indicating that the networks are stable. Therefore, clear within-group weight differences exist between sets of links. A complementary procedure recommended in Epskamp et al., (2018), plotting the values of the link weights and their 95% confidence intervals, also confirmed this conclusion (data not shown).

### *Accuracy of graph-theoretic indices*

In addition to splitting the data into two subgroups, to assess the stability of the results, we followed one of the methods proposed in Epskamp et al. (2018). In brief, a distribution of indices is obtained by bootstrapping, that is, by randomly choosing subsets of subjects. A distribution of correlations across nodes between the original and bootstrapped values of the centrality indices is then calculated. High correlations indicate that the nodes which have the larger centrality values when the whole dataset is used preserve their position when less data are used. The correlation stability coefficient,  $CS(\text{cor}=0.7)$  is defined as the maximum proportion of subjects that can be discarded, such that the correlation between the original and subset indices is 0.7 or higher with 95% probability. Results are deemed accurate if the CS-coefficient is above 0.25. Ideally it should be above 0.5. A third estimation of the replicability of the results was provided by the already described bootstrapped difference test (Epskamp et al., 2018), employed to determine whether the centrality values of different nodes within a diagnostic group were significantly different (data not shown).

The accuracy of the results in Figure 4 was therefore assessed with the correlation stability coefficient, which was calculated for all measures (3) and patient subgroups (6) yielding  $3 \times 6 = 18$  values. 16 of the 18 values of  $CS(\text{corr}=0.7)$  were above 0.5, which indicates that the analysis was very robust in those instances. The two remaining values were below 0.25 and corresponded to the betweenness measures of two of the subgroups, which was consequently the least accurate measure, as the visual inspection of the Figure can confirm. Likewise, Figure A1 also provides information about the accuracy of the results by showing which differences between the indices of different nodes, within subgroup Control A, were statistically significant (black squares) or not (gray squares). Consistent with the previous results, strength and closeness are the measures with more black squares, indicating better discriminability between nodes. Results for the other subgroups were similar.

[INSERT FIGURE A1]

## Tables

Table 1.

*Measures from the National Alzheimer Coordinating Center database.*

Factor	Test	Measure (Label, Maximum Score)
Attention	Digit Span Forward	Longest sequence (DIGFL, 9)
		Total correct Trials (DIGF, 14)
	Digit Span Backward	Longest sequence (DGBL, 8)
		Total correct Trials (DIGB, 12)
Executive Function	Digit Symbol	Total items completed in 90 s (DSYM, 99)
	Part A, Trail Making Test	Total time (TRLA, 150 s)
	Part B, Trail Making Test	Total time (TRLB, 300 s)
Memory	Logical Memory Story A Immediate Recall	Total items recalled (LGMM, 25)
	Logical Memory Story A Delayed Recall	Total items recalled (MMDL, 25)
Language	Animal List generation	Total items in 1 minute (ANIM)
	Vegetable List generation	Total items in 1 minute (VEGT)
	Boston Naming Test	30 odd items (BSTN)

Table 2

*Demographics for the different diagnostic groups.*

Variable	Not Selected	Selected	Control group	MCI group	Dementia group
Age	74.7(10)	73.4(8)	72.7(9)	73.7(8)	75.1(9)
Sex (%Male)	45	44	34	52	54
Race (%White)	83	83	81	85	87
Education (years)	15.6(8)	15.8(6)	16(6)	15.6(6)	15.4(7)
MMSE	20.7(8)	27.2(3)	28.9(1)	26.8(3)	22.3(4)
Participants	3673	15644	7623	5981	2040

*Note:* Numbers indicate mean values with standard deviations in parentheses. MCI=Mild Cognitive Impairment; MMSE=Mini mental state examination. Data for the diagnostic groups correspond to the selected participants.

Table 3

*Descriptive statistics of test measures for the different diagnostic groups.*

Measure	Whole Sample	Controls	MCI	Dementia
DIGF	8.2(2)	8.6(2)	7.9(2)	7.3(2)
DGFL	6.5(1)	6.7(1)	6.4(1)	6(1)
DIGB	6.2(2)	6.8(2)	5.9(2)	4.8(2)
DGBL	4.6(1)	4.9(1)	4.4(1)	3.8(1)
DSYM	40.5(14)	46.8(12)	37.4(12)	26(13)
TRLA	42.8(25)	35(16)	44.8(23)	66.6(38)
TRLB	129.7(82)	93.7(53)	144.1(81)	222.3(87)
LGMM	10.5(5)	13.4(4)	8.9(5)	4.5(4)
MMDL	8.7(6)	12.1(4)	6.7(5)	2.1(3)
ANIM	17.3(6)	20(6)	15.9(5)	11.1(5)
VEGT	12.3(5)	14.7(4)	10.9(4)	7.1(4)
BSTN	25.3(5)	27(3)	24.8(5)	20.4(7)
Number of Subjects	15644	7623	5981	2040

*Note:* Numbers indicate mean values with standard deviations in parentheses. See Table 1 for full name of neuropsychological measures. MCI=Mild cognitive impairment patients; Dementia= Dementia patients.

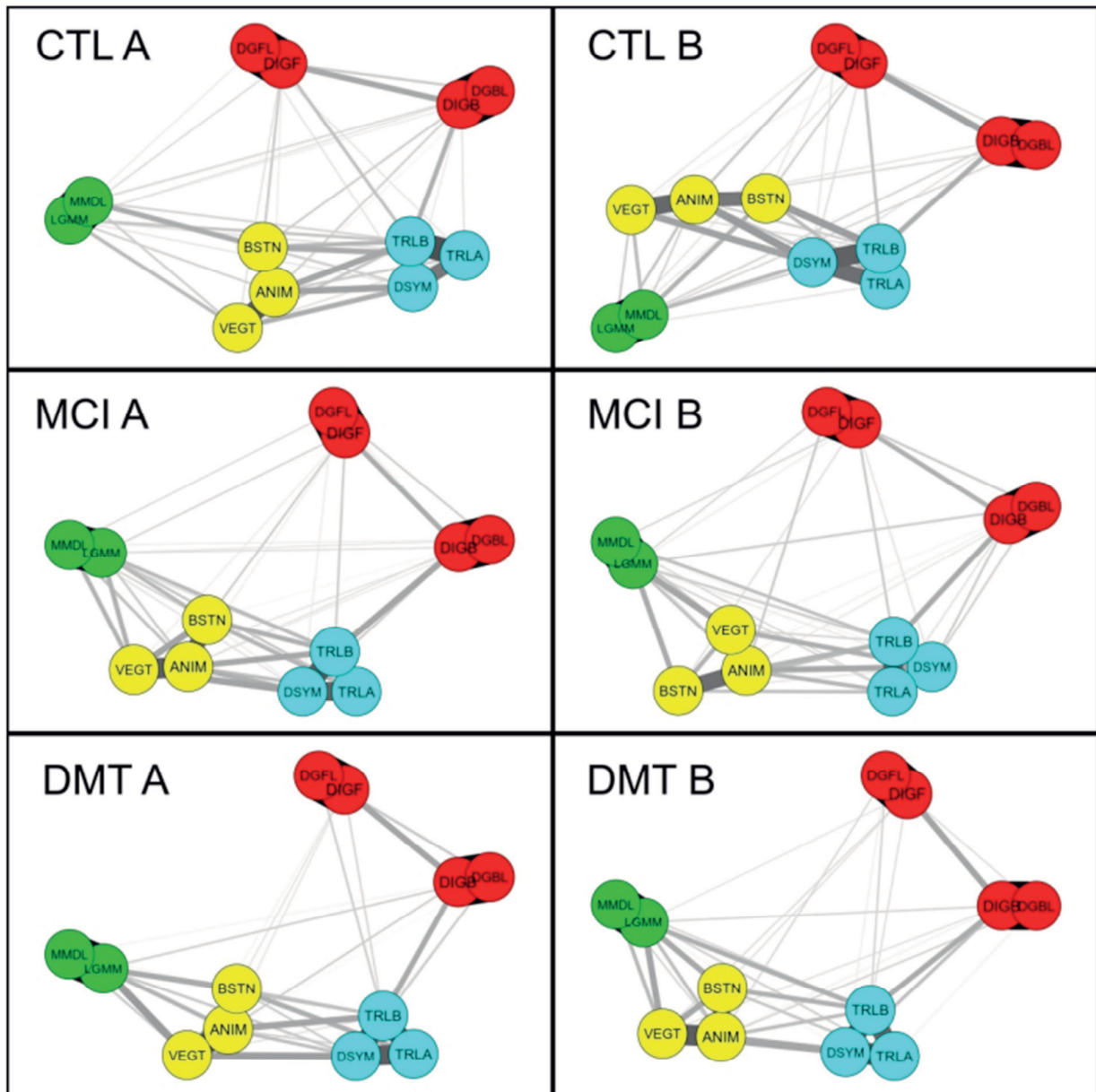


Figure 1. Network graphs for the different diagnostic groups. Nodes colours correspond to red-attention; blue-executive function; yellow-language; green-memory. See Table 1 for full name of neuropsychological tests. CLT=Healthy Controls; MCI=Mild Cognitive Impairment patients; DMT=Dementia patients. A and B indicate data subsampling to check for stability

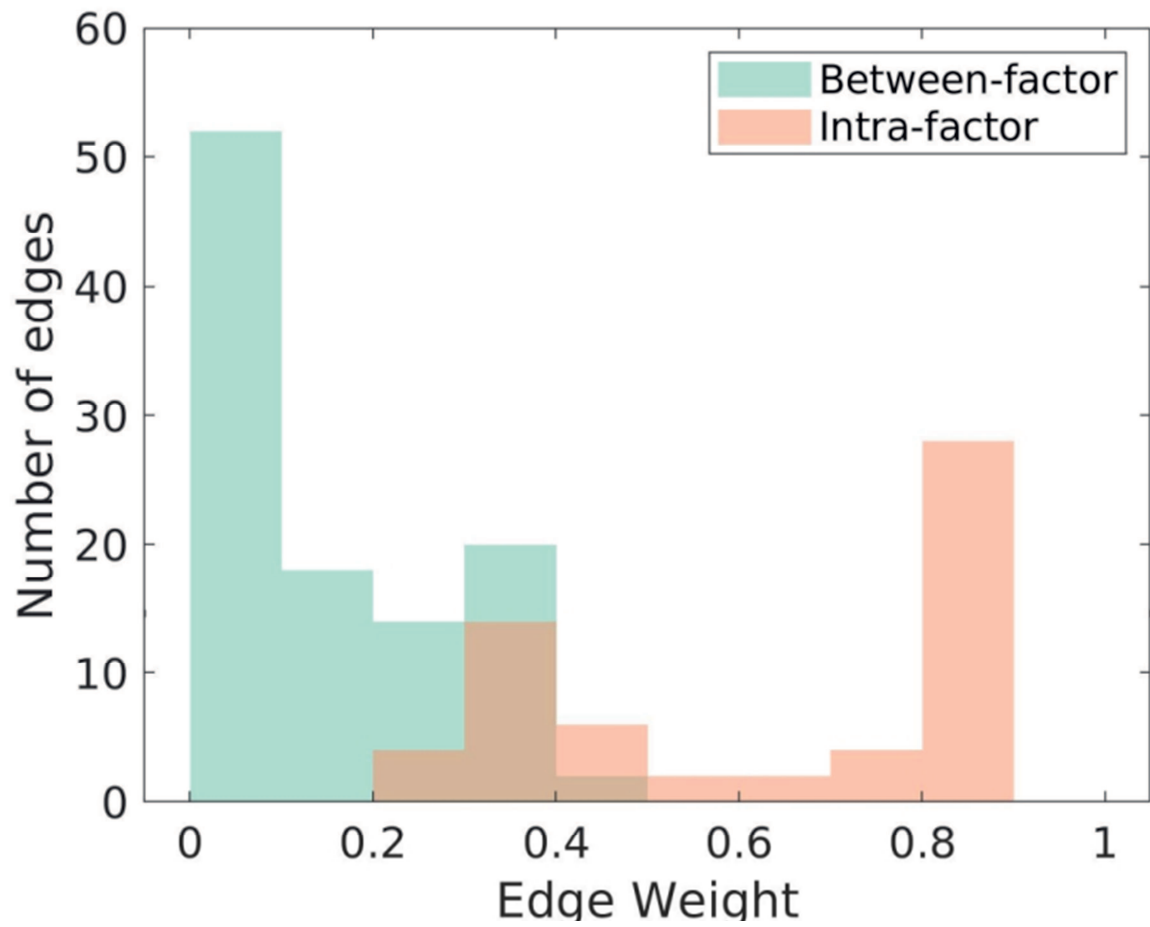


Figure 2. Distribution of link weights from the networks depicted in Figure 1. Between-factor (intra-factor) refers to links connecting nodes corresponding to different factors (the same factor).

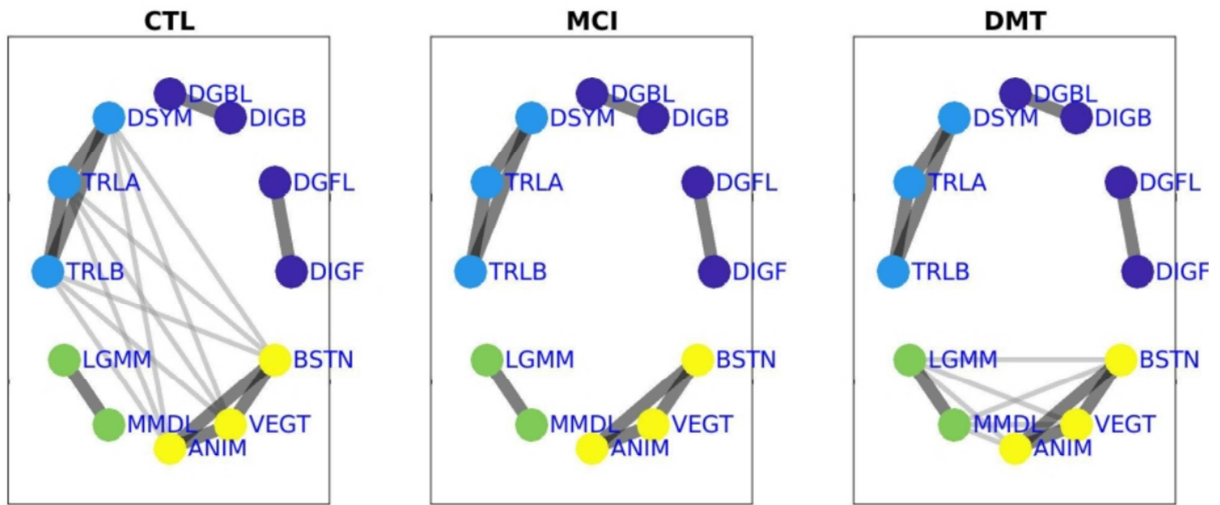


Figure 3. Node classification into communities for the different diagnostic groups. Link widths indicates how often the two connected nodes are classified into the same category. Nodes are colored (Dark blue-attention; blue-executive function; yellow-language; green-memory). CLT=Healthy Controls. MCI=Mild Cognitive Impairment. DMT=Dementia. See Table 1 for full name of neuropsychological tests.

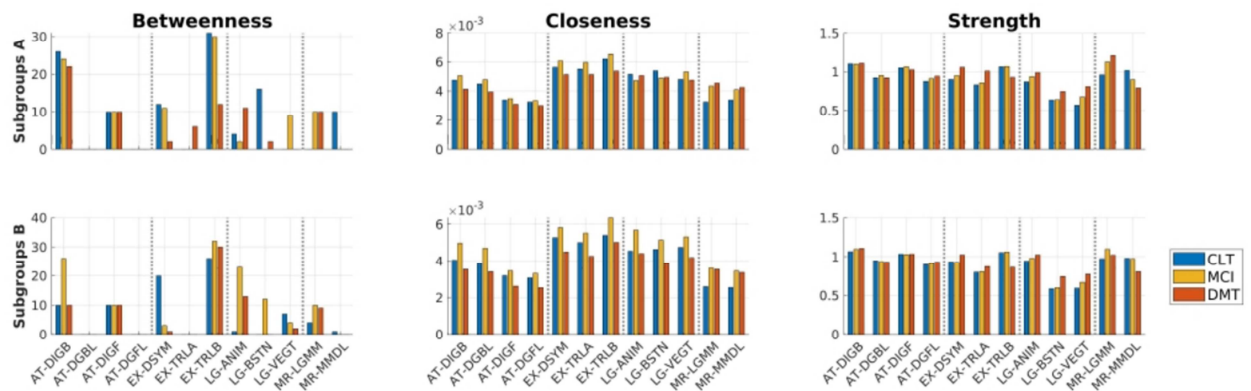


Figure 4. Graph-theory indices (betweenness, closeness and strength) for the different neuropsychological tests. CLT=Healthy Controls. MCI=Mild Cognitive Impairment. DMT=Dementia. A (upper subplot) and B (lower subplot) indicate subsampling. Test labels are preceded by the function they belong to: AT=attention, EX=executive function, LG=language and MR=memory. Test names: DIGFL=digit span forward length, DIGF=digit span forward # correct, DGBL=digit span backwards length, DGBL=digit span backwards length, DIGB=digit span backwards # correct, DSYM=digit symbol, TRLA=trail making A, TRLB=trail making B, LGMM=logical memory immediate, MMDL=logical memory delayed, ANIM=animal list generation, VEGT=vegetable list gen., BSTN=Boston naming.

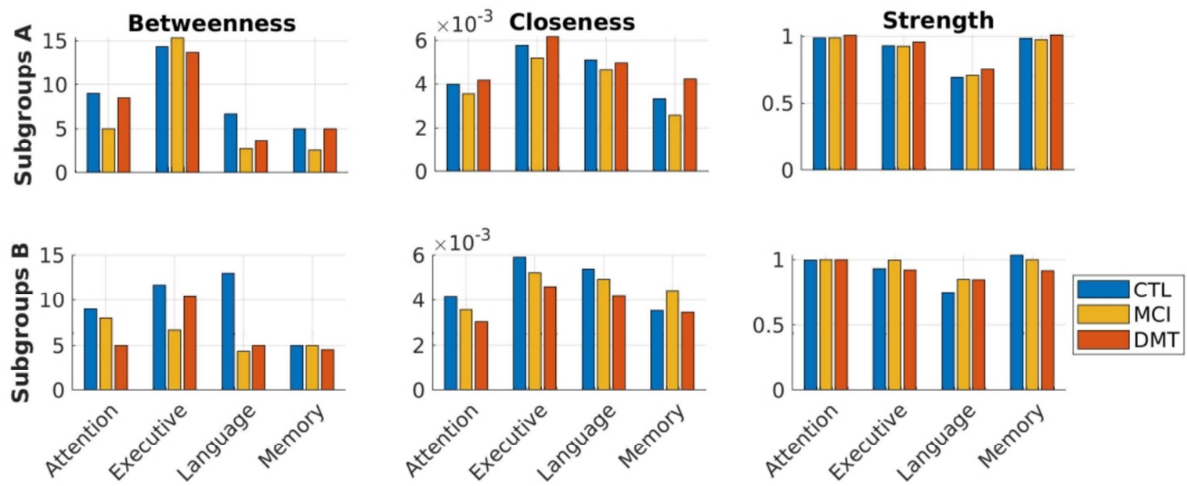


Figure 5. Graph-theory centrality indices averaged according to function. CTL=Healthy Controls. MCI=Mild Cognitive Impairment. DMT=Dementia. The upper subplots correspond to the A subgroups and the lower subplots to the B subgroups

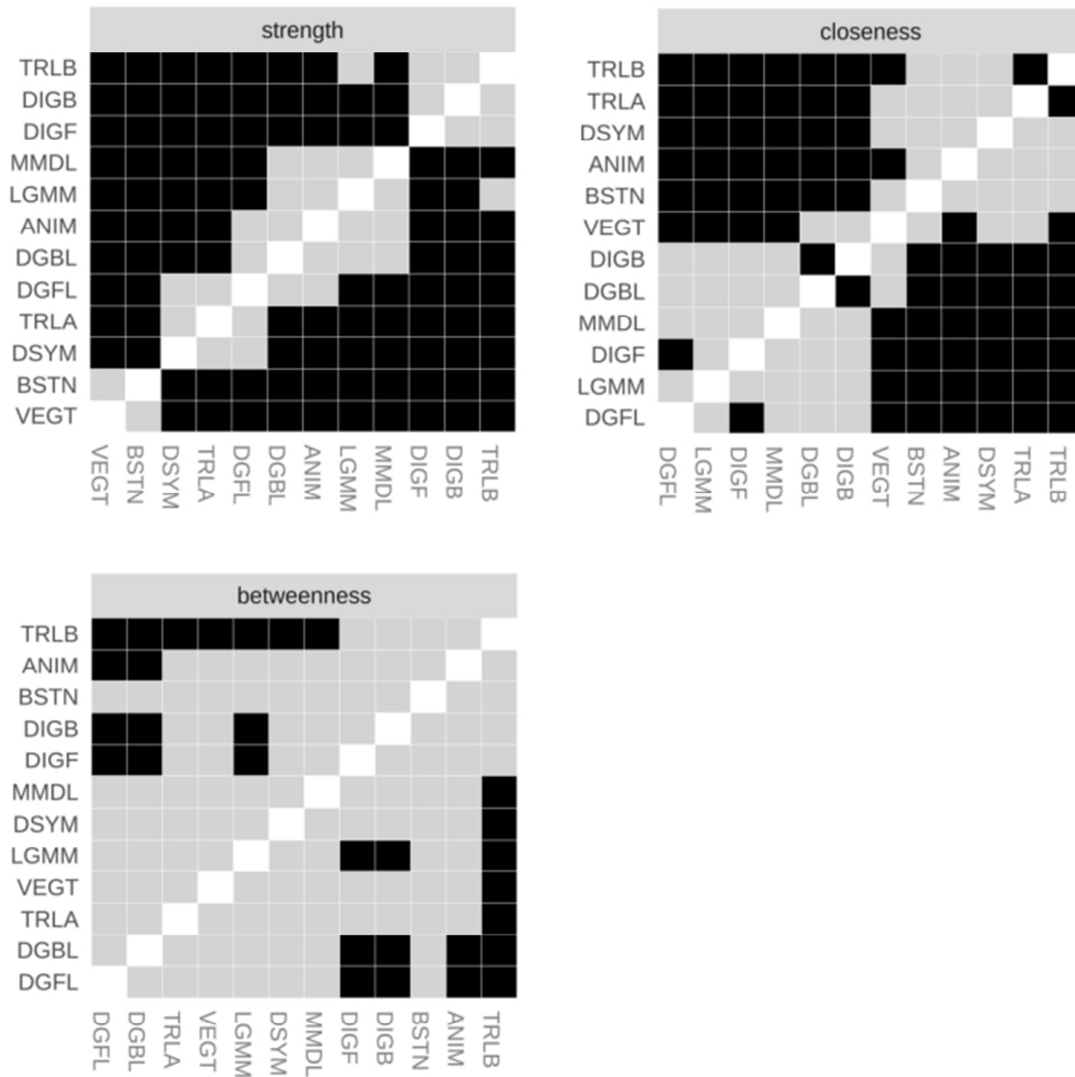


Figure A1. Within-group statistical tests indicating whether the difference in closeness, strength and betweenness, between pairs of nodes is significant (denoted by a black element in the matrix) for subgroup control A.