A network psychometric approach to neurocognitive functioning in Alzheimer's disease

Cameron Ferguson[a]* for the Alzheimer's Disease Neuroimaging Initiative[b]

[a]South Wales Doctoral Programme in Clinical Psychology, Cardiff University, Wales; [b]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf
* fergusonc3@cardiff.ac.uk; https://orcid.org/0000-0001-5248-1182; https://twitter.com/C_E_Ferguson; https://www.linkedin.com/in/cameron-ferguson-28b8ab121/

Published: Ferguson, C. (2021). A network psychometric approach to neurocognition in early Alzheimer's disease. Cortex, 137, 66 – 73. doi.org/10.1016/j.cortex.2021.01.002.

Abstract

In a typical pattern of Alzheimer's disease onset, episodic memory decline is predominant while decline in other neurocognitive domains is subsidiary or absent. Such descriptions refer to relationships between neurocognitive domains as well as deficits within domains. However, the former relationships are rarely statistically modelled. This study used psychometric network analysis to model relationships between neurocognitive variables in cognitive normality (CN), amnestic mild cognitive impairment (aMCI), and early Alzheimer's disease (eAD). Gaussian graphical models with extended Bayesian information criterion graphical lasso model selection and regularisation were used to estimate network models of neurocognitive and demographic variables in CN (n = 229), aMCI (n = 395), and eAD (n = 191) groups. The edge density, network strength and structure, centrality, and individual links of the network models were explored. Results indicated that while global strength did not differ, network structures differed across CN and eAD and aMCI and eAD groups, suggesting neurocognitive reorganisation across the eAD continuum. Episodic memory variables were most central (i.e., influential) in the aMCI network model, whereas processing speed and fluency variables were most central in the eAD network model. Additionally, putative clusters of memory, language and semantic variables, and attention, processing speed and working memory variables arose in the models for the clinical groups. This exploratory study shows how psychometric network analysis can be used to model the relationships between neurocognitive variables across the eAD continuum and to generate hypotheses for future (dis)confirmatory research.

Keywords: Alzheimer's disease; neurocognition; neuropsychology; network psychometrics; Gaussian graphical model

1. Introduction

In a typical pattern of Alzheimer's disease (AD) onset, predominant episodic memory decline is observed before decline in other neurocognitive domains becomes apparent with disease progression (Howieson et al., 2008; Kolb, 2015; Lezak, Howieson, Bigler, & Tranel, 2012; Rushing, Sachs-Ericsson, & Steffens, 2014; Weintraub, Wicklund, & Salmon, 2012). Neuropsychologists can be relatively confident that it is memory that is affected, given disproportionately poor performance on a memory test, in part, because of latent variable models. Latent variable models model the shared variance underlying task performance, which can be taken to signify neurocognitive domains. This type of statistical modelling can be useful because neuropsychology operates in a reality in which the lack of test-function specificity is common (Lezak et al., 2012).

Nevertheless, descriptions of the typical pattern of decline in AD refer to the relationships between neurocognitive domains as well as performance on measures of each domain. Episodic memory decline is predominant, while decline in other neurocognitive domains is subsidiary or not yet present (Howieson et al., 2008; Kolb, 2015; Lezak et al., 2012; Rushing et al., 2014; Weintraub et al., 2012). Statistically, these descriptions invoke the unique variance between variables as well as shared variance signifying domains (Costantini et al., 2015; Epskamp & Fried, 2018). The latent-variable approach, with an emphasis on common variance and measurement invariance of neurocognitive domains (Bowden, Cook, Bardenhagen, Shores, & Carstairs, 2004; Meredith, 1993; Strauss & Smith, 2009), is not best placed to model such between-variable relationships across the early AD (eAD) continuum. However, it is important

that these relationships are calculated as robustly and described as precisely as possible. Indeed, the meaningfulness of much neuropsychological data emerges from consideration of the relationships amongst variables, such as the presence or absence of deficits in some functions relative to others and the influence (whether compensatory or detrimental) of one function on another.

The network psychometric approach is well placed to model the relationships between neurocognitive variables in eAD, while complementing the more traditional latent variable approach. In essence, network psychometric models represent relationships amongst variables once relationships between all other variables in the model have been accounted for. The remaining associations, or 'edges' in network terminology, between variables, or 'nodes', can then be visualised as a network structure in which the nodes are assumed to influence each other. Analysis of this network structure, both visually and through network metrics, can offer useful insights into relationships between variables. Moreover, network estimation gives rise to mathematically precise models of these relationships (van Bork, van Borkulo, Waldorp, Cramer, & Borsboom, 2018), which can help to quantify verbal descriptions of typical relationships amongst variables.

One potentially illuminating network metric is strength centrality. Centrality refers to the importance of a node in forming the network structure, an implication being that not all nodes are equally influential in the network (Costantini et al., 2015). Strength centrality is simply the sum of the edge weights that are connected to a node (van Bork et al., 2018). A node that has high strength in a network model is thought to directly influence many other nodes in the network, without mediation from other nodes (Costantini et al., 2015). The description of the typical neuropsychological profile in early AD does not necessarily imply that memory nodes, or indeed any nodes, are the most central in the network, although it may be that memory variables are most central. However, it is fruitful to investigate the strength centrality properties of network models as knowledge of the most central nodes can help to predict the ability of the network to compensate for deficits in a particular node and/or identify targets for interventions (Costantini et al., 2015). Accordingly, the primary aim of this study was to statistically model network structure and calculate strength centrality indices, and other network properties, at three points along the eAD continuum: cognitive normality (CN), amnestic mild cognitive impairment (aMCI), and eAD.

Only one study has used the network psychometric approach to study neurocognition in AD. Tosi et al. (2020) used network psychometrics to model the performance of older adults in CN, vascular encephalopathy (VE), and AD groups on neuropsychological batteries sampling a range of neurocognitive domains. The authors also included data from screening tools and demographic variables (age, sex, education) in their network models, enabling consideration of the impact of these factors on neuropsychological test performance. Network estimation generated a relatively sparse network model for the CN group. By contrast, a denser structure containing two clusters of nodes was observed in the AD network model, indicating reorganisation of the relationships amongst neurocognitive variables in the disease state. One cluster consisted of tests related to memory and the other comprised tests with an executive component. Category fluency was the most central node in this network model, with high strength and betweenness values. It also served as a bridge between the memory and executive clusters. The authors also generated a number of hypotheses on the basis of their models, which highlights the utility of network psychometrics as an exploratory

tool for generating hypotheses for future confirmatory research (Epskamp & Fried, 2018).

Tosi et al. (2020) analysis is a promising step forwards in the application of network psychometrics to the neuropsychology of healthy aging and dementia. However, their study presents two avenues for improvement. First, different variables were included in each of the network models. These discrepancies limit the comparability of the models, even though the different tests probe the same neurocognitive domains, because different variables may introduce different patterns of dependencies in network models. Second, they did not quantitatively compare the networks, which would have been inappropriate given the different number of variables in the healthy and clinical network models. Quantitative comparison of network models is complementary to visual analysis and may overcome some of the limitations inherent in eyeballing data. The present study sought to improve on these limitations by quantitatively comparing network strength and structure across groups. Additionally, the study devoted much needed attention to the eAD spectrum, where interventions based on network approaches to neurocognition are likely to be most fruitful.

In summary, relationships between neurocognitive variables are often referred to in descriptions of AD onset. However, these relationships have not been statistically modelled at multiple points across the eAD continuum. Psychometric network modelling facilitates the analysis of the unique variance between variables in a complementary manner to latent variable modelling. The present study aimed to estimate and compare the properties of neurocognitive network models for three groups along the eAD continuum, and to use these network models to generate hypotheses for future (dis) confirmatory research.

2. Method and materials

The following sections describe how the data were sourced, all inclusion/exclusion criteria, and all variables in the study. There were no experimental manipulations. All R code is accessible online: https://osf.io/uzjem/. No part of the study was preregistered in a public forum. The proposed analyses were approved by the Alzheimer's Disease Neuroimaging Initiative (ADNI) Data and Publications Committee.

2.1. Dataset and pre-processing

Data used in this study were sourced from the ADNI database (http://adni.loni.usc.edu/data-samples/access-data/), which has been described in detail elsewhere (Petersen et al., 2010). The ADNI is led by Michael W. Weiner, MD, and up-to-date information on the initiative can be accessed at www.adniinfo.org/. For this study, all ADNI participants who participated in baseline neuropsychological assessment were eligible for inclusion. Participants were excluded if they were missing diagnostic data, demographic data, or more than 10% of their neuropsychological test data. Mean imputation was used for cases with less than 10% missing data.

2.2. Participants

Data from a total of 815 ADNI participants were selected for analysis. Demographic information and dementia severity for each group is presented in Table 1. Participants were grouped into CN, aMCI, and eAD groups on the basis of the diagnosis (or lack of) they received during their initial ADNI visit.

The ADNI used the Petersen et al. (2010) diagnostic criteria for amnestic MCI and the National Institute of Neurological and Communicative Disorders and Stroke -

Alzheimer's Disease Related Disorders Association criteria for probable AD (McKhann et al., 1984). ADNI exclusion criteria included a Hachinski Ischemic Score (Moroney et al., 1997; Rosen, Terry, Fuld, Katzman, & Peck, 1980) greater than 4; a Geriatric Depression Scale (Yesavage et al., 1983) of 6 or higher; inadequate visual and auditory acuity for neuropsychological testing; changes in medication for 4 weeks prior; and poor general health (Petersen et al., 2010). Accordingly, vascular or mixed dementia, depression and other confounding factors were unlikely to influence the neuropsychological data.

Dementia severity was measured by the Clinical Dementia Rating Scale (CDR), which is a 5-point clinician-rated scale of memory, orientation, judgement and problem solving, community affairs, functioning regarding at home and with hobbies, and personal care (Morris, 1993). A score of 0 signifies the absence of dementia and higher scores represent more severe dementia. The Mini-Mental State Exam (MMSE) was also used (Folstein, Folstein, & McHugh, 1975). AD participants (CDR mean = .7, SD = .3; MMSE mean = 23.1, SD = 2.1) were therefore considered to have mild or eAD according to the CDR and MMSE.

Table 1. Demographic data and dementia severity by group

	Group		
	Cognitively normal	Amnestic mild cognitive impairment	Early Alzheimer's disease
n	229	395	191
Gender			
Male (%)	52%	64.3%	52.4%
Female (%)	48%	35.7%	47.6%
Age (years)	76.0 (5.0)	74.9 (7.4)	75.4 (7.5)
Education (years)	16.1 (2.9)	15.7 (3.0)	14.8 (3.0)
Premorbid ability			
ANART (errors)	9.1 (8.0)	13.6 (9.9)	15.8 (10.0)
Dementia severity			
CDR	0.0 (0.0)	0.5 (0.0)	0.7 (0.3)
MMSE	29.1 (1.0)	27.0 (1.8)	23.4 (2.1)

NB. brackets contain standard deviations. ANART: American National Adult Reading Test (Blair & Spreen, 1989). CDR: Clinical Dementia Rating Scale (Morris, 1993). MMSE: Mini-Mental State Exam (Folstein et al., 1975).

2.3. Neurocognitive variables

Neuropsychological tests were selected from the battery administered to ADNI participants. Variables with restricted variance, such as error scores on the trail making tests and category fluency, were excluded. This was necessary given the network modelling procedures, which depended on the assumption of normality (Epskamp & Fried, 2018), and the between subjects approach of the study, which required adequate variance to produce reliable results (Hedge, Powell, & Sumner, 2018). For the eligible variables, non-paranormal transformation was used on raw scores to approximate the normal distribution (Liu, Lafferty, & Wasserman, 2009). Descriptive statistics are presented in Table 2.

2.3.1. Attention/processing speed

2.3.1.1. Trail Making Test Part A and B.
Trail Making Test Part A (TMTA) requires the participant to draw lines to connect a scrambled array of the numbers 1 to 25 in ascending order (Reitan, 1958). Trail Making Test Part B (TMTB) requires the participant to perform an ostensibly similar task while alternating between the numbers 1 to 13 and the letters A to L in ascending order (i.e., 1-A-2-B-3-C, etc.) To make partial correlations in the network models easier to interpret, completion times for TMTA and TMTB were reversed (i.e., signed negatively) so that higher scores reflected better performance.

2.3.1.2. Wechsler Adult Intelligence Scale - Revised Digit Symbol Substitution.
Digit-Symbol Substitution (DSS) requires the participant to match symbols to digits according to an explicit key under a strict 90 sec time limit (Wechsler, 1981). The DSS score corresponds to the total number of correct responses generated within the time limit.

Table 2. Descriptive statistics

		Group					
		Cognitively Normal		Mild Cognitive Impairment		Alzheimer's Disease	
Domain	Test	*M*	*SD*	*M*	*SD*	*M*	*SD*
Attention	DSS	45.7	10.2	36.8	11.2	26.5	13.3
	TMT						
	Part A	36.4	13.2	44.5	22.2	66.9	37.0
	Part B	89.2	44.3	129.4	74.0	186.7	95.6
Working memory	Digit Span						
	Forwards	8.8	2.0	8.3	2.0	7.6	1.9
	Backwards	7.2	2.2	6.1	2.1	4.9	2.0
Episodic memory	AVLT						
	Immediate	43.0	9.8	30.7	9.0	23.1	7.8
	Intrusions	1.5	1.8	1.4	1.6	1.0	1.9
	Delayed	7.4	3.7	2.8	3.3	0.7	1.6
	Recognition	12.8	2.7	9.7	3.6	7.2	4.0
	Recognition Errors	0.8	1.2	3.0	2.2	1.9	2.3
Language	BNT (30 items)	27.3	3.2	25.0	4.4	20.3	21.9
Fluency	Category fluency	34.6	8.1	26.6	7.3	20.2	7.4
Visuospatial	Clock drawing	4.7	0.7	4.2	1.0	3.4	1.3
	Clock copying	4.8	0.6	4.6	0.7	4.3	1.0
Premorbid ability	ANART	9.1	8.0	13.6	9.9	15.8	10.0

DSS: Digit Symbol Substitution. TMT: Trail Making Test. AVLT: Rey Auditory Verbal Learning Test. BNT: Boston Naming Test. ANART: American National Adult Reading Test.

2.3.2. Working memory

2.3.2.1. Wechsler Memory Scale - Revised Digit Span Forwards and Backwards.

Digit span forwards (DSF) requires the participant to attend to and repeat a string of single digits in the order in which it was heard, with each string length increasing by 1 digit on subsequent trials. Digit span backwards (DSB) requires the participant to repeat each increasingly long string in the reverse order to which it was heard (Wechsler, 1987). The DSF and DSB scores represent the total number of trials completed correctly.

2.3.3. Memory

2.3.3.1. Rey Auditory Verbal Learning Test.

During the Rey Auditory Verbal Learning Test (AVLT), the participant attempts to memorise a list of 15 semantically unrelated words over 5 trials, each of which ends in a free-recall trial. Following these 5 trials, a new list of 15 unrelated 'distractor' words is read to the participant, and they are asked to repeat the original word list. The participant is asked to repeat the original word list again after a 30-minute delay period. In the ADNI, AVLT Immediate was the sum of correct responses on trials 1 to 5 (Rey, 1964). AVLT Intrusions was the sum of intrusion errors on trials 1 to 5. AVLT Delayed was the number of words from the original word list repeated after the 30minute delay. AVLT Recognition was the number of correctly identified target words on a recognition trial in which all of the words from the original and 'distractor' lists plus 20 phonemically and/or semantically similar words (50 words total) were presented to the participant. AVLT Recognition Errors was the number of incorrect responses on the recognition trial. Intrusions and Recognition Errors were resigned negatively so that lower scores signify poorer performance.

2.3.4. Language

2.3.4.1. Boston Naming Test.

The Boston Naming Test (BNT) is a confrontation naming task in which the participant is presented with 60 sequential line drawings of increasingly difficult to identify items and asked to name them (Kaplan, Goodglass, & Weintraub, 1983). In the ADNI, participants were shown 30 stimuli. The BNT score represents the total number of spontaneously provided correct answers, without reliance on semantic or phonemic cues.

2.3.5. Visuospatial

2.3.5.1. Clock Drawing and Copying. The clock drawing task involves drawing a clock, without the aid of a visual example. Clock copying involves drawing a clock with a visual example present (Goodglass & Kaplan, 1983). The ADNI used a 5-point scale to credit participants for an approximately circular face, symmetry of number placement, correctness of numbers, presence of two hands and the presence of two hands set to the specified time.

2.3.6. Fluency

2.3.6.1. Category Fluency.

In category fluency tasks, the participant must name as many instances of a semantic category, in this case animals and vegetables, as they can within a 60 second time limit (Morris et al., 1989). In the present study, total correct scores were summed prior to

transformation to create a single category fluency variable. The score reflects the total number of correct responses for both categories.

2.4. Network estimation

Three Gaussian graphical models (GGM) with extended Bayesian information criterion graphical lasso (EBICglasso) model selection and regularisation were estimated on the basis of the neurocognitive and demographic data for the CN, aMCI, and eAD groups.

2.4.1. The Gaussian graphical model

The GGM is an undirected graphical model, or Markov random field. The GGM is computed on the inverse of a covariance matrix, which encodes partial correlations (Epskamp, Borsboom, & Fried, 2018). Subsequently, in a GGM, edges can be interpreted as partial correlation coefficients ranging from -1 to 1 (Epskamp, Borsboom, et al., 2018; Epskamp, Waldorp, Mottus, & Borsboom, 2018; Lauritzen, 1996); although, these will differ from standard partial correlation coefficients due to the model selection and regularisation procedures employed. The edges represent the remaining association between two variables after conditioning on all other variables. All GGMs were estimated in R using the 'qgraph' package (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012) as implemented in the 'bootnet' package version 1.3 (Epskamp, Borsboom, et al., 2018).

2.4.2. Model selection and regularisation

Network models are subject to sampling variation, which can introduce unreliable associations (Epskamp & Fried, 2018). To promote replicability and guard against overinterpretation of network models it is important to limit the number of spurious edges in a model. EBICglasso model selection and regularisation was employed as it returns a sparse network structure (i.e., one with relatively few edges) in comparison to less conservative model selection techniques. As the study was exploratory, emphasising specificity at the potential cost of sensitivity was preferential to lessen the chance that hypotheses generated by analysing the network models might be due to false positive edges. The least absolute shrinkage and selection operator (lasso) works by constraining the sum of partial correlation coefficients, meaning that all edge weight estimates are pulled towards zero and some are estimated as zero (Tibshirani, 1996). In EBICglasso, the lasso tuning parameter is assigned automatically by minimising the extended Bayesian information criterion (EBIC). An EBIC hyperparameter determines how much the criterion prefers simpler models. An EBIC hyperparameter of 0 favours sensitivity, and runs the risk of including spurious edges, while a setting of 0.5 privileges specificity, and runs the risk of reducing sensitivity when the true network structure is dense (Epskamp & Fried, 2018). The EBIC hyperparameter value used in the study (0.25) was selected by optimising the desired psychometric properties, which were estimated with simulation studies.

2.4.3. Simulation studies

Three simulation studies were conducted to investigate the impact of various model selection parameters and sample sizes on the psychometric properties of the network models, and to select optimal parameters for promoting specificity. The psychometric properties of interest were the sensitivity, specificity and correlation of the network model to a large number of comparator models. Simulation studies were performed under each observed network model (CN, aMCI, and eAD) for varying sample sizes and

EBIC hyperparameter values. They can be described in terms of a five-step process. First, random data was generated under the parameters of an observed network model, which acted as the input. Second, while the parameters (E.g., edge weights, network structure, etc.) of the input network model were held constant, different values were set for the EBIC hyperparameter (0, 0.25, 0.5) and sample size (100, 200, 500). Third, a comparator network model was estimated on the basis of the randomly generated data from the input network model and the various altered factors. Fourth, the comparator network was compared to the input network model. Fifth, stages two to four were repeated 1000 times, which gave rise to estimates of the sensitivity, specificity, and correlation between comparator and input networks under different conditions. Simulation studies indicated that an EBIC hyperparameter value of 0.25 and the observed sample sizes were sufficient for ensuring specificity without negating sensitivity. The simulation studies were also conducted within the 'bootnet' package (Epskamp, Borsboom, et al., 2018). The results of the simulation studies are available in Appendix A.

2.4.4. Post-hoc investigations

2.4.4.1. Edge weight stability

Because network models are subject to sampling variation, it is also important to investigate the stability of the edge weight estimates in the final network models to ensure that undue certainty is not applied when interpreting them. Non-parametric bootstrapping was employed to generate 95% confidence intervals around edge weights in each group using a five-step process, also implemented in 'bootnet' (Epskamp, Borsboom, et al., 2018). First, edge weights were computed in the original sample for each group. Second, a new comparator dataset was generated by randomly sampling from the original dataset with replacement. Third, edge weights were computed in the new comparator dataset. Fourth, steps two (resampling from data with replacement) and three (computing statistics in generated data) were repeated 1000 times. Finally, the ranges of the computed edge weights were used to draw confidence intervals for each edge weight in each network model. The full results of these analyses are presented in Appendix B.

2.4.4.2. Edge density

The edge density index was used to describe the density of each network model. This descriptive statistic is calculated by simply dividing the number of observed edges by the number of possible edges. The number of possible edges is calculated first using the simple formula $P(P-1)/2$ in which P denotes the number of nodes in the network model (Epskamp et al., 2017). The number of possible edges for each model was 136, given the presence of 17 nodes.

2.4.4.3. Comparison of network structures

The network comparison test (NCT) was used to explore the invariance of network structure and global strength across CN, aMCI, and eAD neurocognitive network models. 'Network structure' refers to the assumption that the network as a whole (i.e., the specific patterns of edges connecting nodes) is identical across groups. 'Global strength' refers to the assumption that the absolute sum of all edges is the same across networks (Opsahl et al., 2010; van Borkulo et al., 2017). The NCT is a non-parametric permutation-based hypothesis test which employs five stages to determine whether certain statistics from two groups differ (van Borkulo et al., 2017). First, the NCT

computes the statistic of interest (e.g., the network structure) of two network models. Second, all cases (i.e., participants from sample A and sample B) are pooled in a single, larger dataset. Third, two new groups are created by randomly redistributing the cases (i.e., cases from different initial groups are highly likely to be combined within the two new groups). Fourth, the statistic (e.g., the network structure) is computed for the two new groups and steps two to four are repeated 1000 times to obtain a null distribution. Finally, the NCT tests whether the observed difference computed in the original group is in the null distribution arising from step four. If the original statistic is in the null distribution, then the networks do not differ with regards to that statistic. In other words, the NCT probes whether observed differences in a statistic between both groups is larger than might be expected from a null distribution in which these groups come from the same population. The p values generated by the NCT are equal to the proportion of test statistics in the null distribution that are at least as extreme as the observed test statistic. As the study was exploratory, a threshold was not set for considering NCT-generated p values statistically significant or not. Rather, p values for each analysis were interpreted as rough indices of surprisingness given the assumption that the observed statistics did not differ across groups.

2.4.4.4. Centrality indices

Node strength and betweenness were calculated for each network model and reported as standardized z scores. Before these centrality estimates were interpreted, their stability was probed with case-dropping bootstrapping. These bootstraps worked in a similar way to the bootstrapping procedures described above; however, they used subsampling without replacement and diminishing sample sizes rather than resampling with replacement and constant sample sizes. These methods generated a centrality stability coefficient (CS), which quantified, with 95% confidence, the greatest proportion of participants that could be dropped to preserve a correlation of higher than 0.7 with the original centrality values. A CS of 0.5 or higher is indicative of a stable centrality estimate across diminishing cases and a CS that does not exceed 0.25 should not be interpreted (Epskamp & Fried, 2018). Only the strength index met the minimum criterion for stability; therefore, betweenness centrality estimates are not presented or interpreted in the paper. The full results of the strength centrality analyses are available in Appendix C.

3. Results/discussion

Network plots for the CN, aMCI, and eAD groups are presented in Figures 1, 2, and 3, respectively.

3.1. Cognitive normality

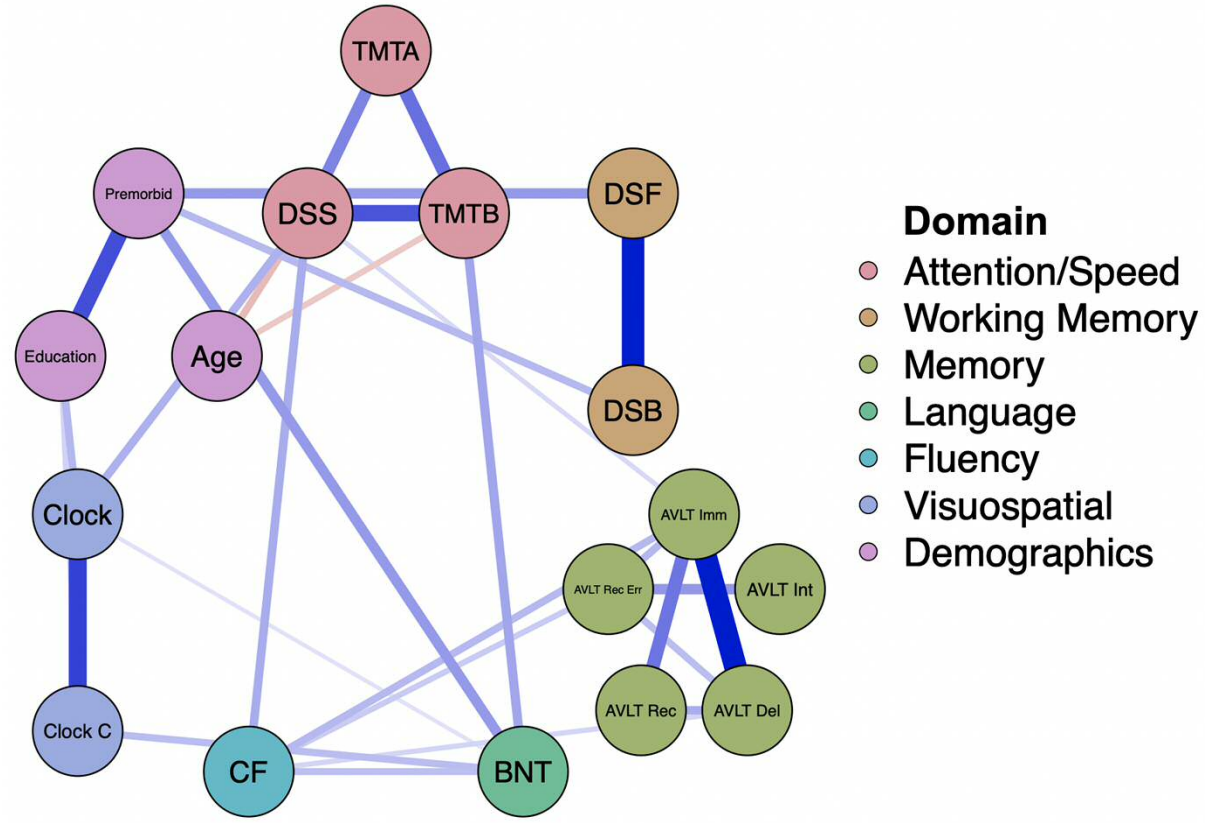
The CN network model, shown in Figure 1, was relatively dense (edge density = 0.79), with multiple links between tests of different neurocognitive domains, although some of these edges were relatively weak (e.g., DSS-AVLT Immediate, bootstrapped mean edge weight = 0.04, 95% CI = 0.02, 0.13).

The most central nodes in the model were AVLT Immediate (strength centrality z score = 1.99) and DSS (z = 1.28), both of which were linked to various tests of different domains, indicating that memory acquisition and attentional/processing speed abilities were particularly important influences on neurocognitive functioning for the CN older adult group. Premorbid ability (ANART) was also highly central (z = 1.11), consistent with the lack of pathology in the CN group and the well-established association between

general cognitive ability and functioning in neurocognitive domains (Binder et al., 2009; Diaz-Asper, Schretlen, & Pearlson, 2004; Mohn, Sundet, & Rund, 2014).

The strongest edges in the network model were between tests representing the same neurocognitive domain. Additionally, these more substantial edges were present amongst all tests of the same neurocognitive domain. These two findings highlight the shared variance associated with neurocognitive domains and suggest that they provide a good account of neurocognitive functioning for a group of healthy older adults.

Figure 1. Neurocognitive network model for cognitively normal older adults



DSS: Digit Symbol Substitution. TMTA: Trail Making Test Part A. TMTB: Trail Making Test Part B. DSF: Digit Span Forward. DSB: Digit Span Backward. AVLT Imm: Rey Auditory Verbal Learning Test Immediate. AVLT Imm: Rey Auditory Verbal Learning Test Intrusions. AVLT Del: Rey Auditory Verbal Learning Test Delayed. AVLT Rec: Rey Auditory Verbal Learning Test Recognition. AVLT Rec Err: Rey Auditory Verbal Learning Test Recognition Errors. BNT: Boston Naming Test. CF: category fluency. Clock: clock drawing. Clock C: clock copying. Education: education (years). Age: age (years). Premorbid: American Adult National Reading Test.

In some cases, edges between tests of different domains were relatively substantial. For example, the edge between AVLT Immediate and category fluency (bootstrapped mean edge weight = .15, 95% CI = .04, .26), after conditioning on all other variables in the network. This edge could reflect the co-dependence of semantic and episodic memory during episodic memory encoding (Greenberg & Verfaellie, 2010), at least for list-learning tasks which feature semantically unorganised information.

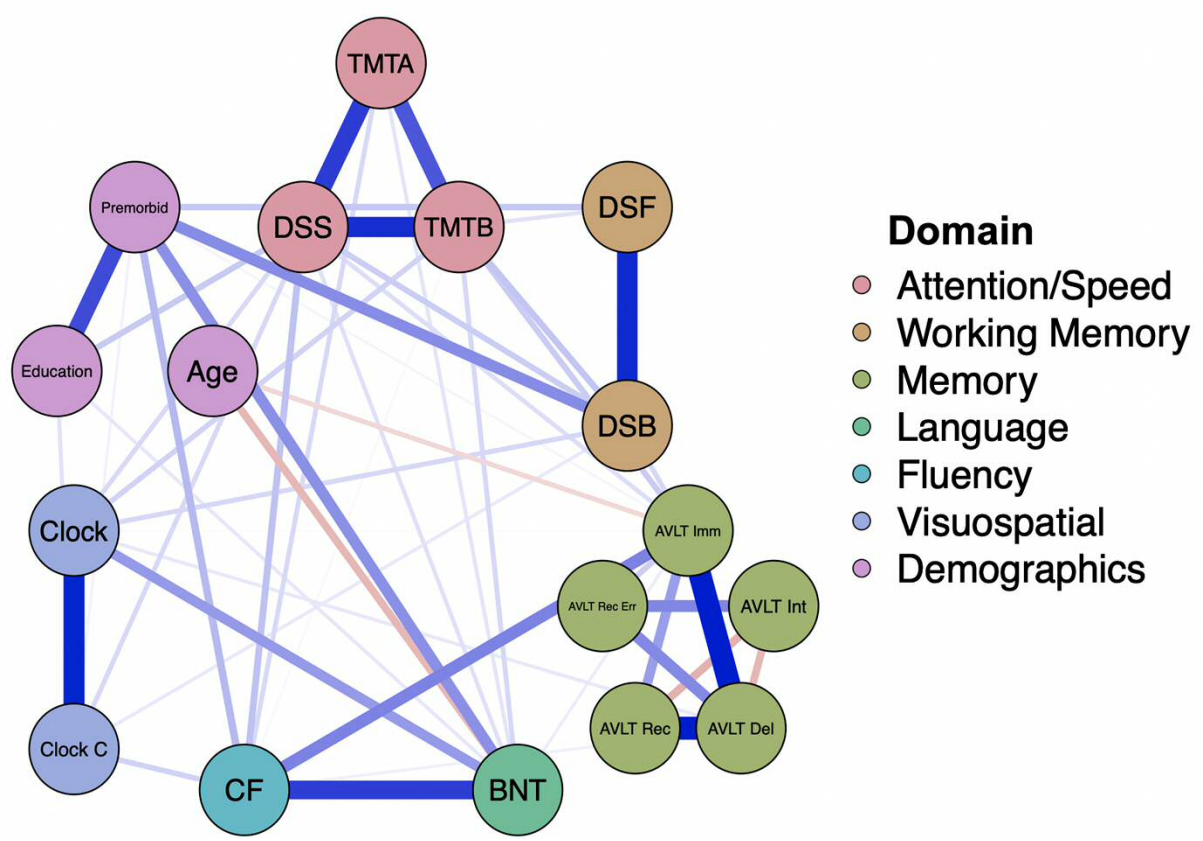
Together, the properties of the CN network model suggest that neurocognitive variables had multiple small influences on each other, irrespective of their purported domain; however, neurocognitive functioning was largely characterised by the presence

of fractionated neurocognitive domains. Network qualities were different in the aMCI and eAD models, suggesting neurocognitive reorganisation in the disease states.

3.2. Amnestic mild cognitive impairment

As shown in Fig. 2, the aMCI network model was also relatively dense (edge density = .78). In comparison with the CN model, the global strength of the aMCI model was not surprisingly different (NCT global strength comparison: p = .47, test statistic = .44). Neither was the overall structure of model surprisingly different from the CN model (NCT network structure comparison: p = .18, test statistic = .23). Descriptively, there were more edges linking tests of different domains in the aMCI model than in the CN model, which suggests some reorganisation of the relationships amongst neurocognitive variables in the disease state. For the most part, these edges were weaker than those connecting tests of the same domain. This tendency suggests that the neurocognitive status of the aMCI group was still partly characterised by the presence of relatively distinct neurocognitive domains.

Figure 2. Neurocognitive network model for older adults with amnestic mild cognitive impairment



DSS: Digit Symbol Substitution. TMTA: Trail Making Test Part A. TMTB: Trail Making Test Part B. DSF: Digit Span Forward. DSB: Digit Span Backward. AVLT Imm: Rey Auditory Verbal Learning Test Immediate. AVLT Imm: Rey Auditory Verbal Learning Test Intrusions. AVLT Del: Rey Auditory Verbal Learning Test Delayed. AVLT Rec: Rey Auditory Verbal Learning Test Recognition. AVLT Rec Err: Rey Auditory Verbal Learning Test Recognition Errors. BNT: Boston Naming Test. CF: category fluency. Clock: clock drawing. Clock C: clock copying. Education: education (years). Age: age (years). Premorbid: American Adult National Reading Test.

While global network strength and structure did not appear to differ across the aMCI and CN models, and weakened neurocognitive domains were still visible in the aMCI model, two putative clusters emerged. One of these consisted of tests probing memory, semantic, and language abilities (namely the AVLT, BNT, and category fluency); the other was formed of tests with attention, processing speed, and working memory components (including DSS and the digit span and trail making tests). This pattern of association differed from the more fractionated neurocognitive domains shown in the CN network model, which were less distinguishable in the aMCI network. The constellations of edges amongst tests of memory-semantic-language abilities and tests with attention-speed-working memory components in the aMCI model may reflect the divergence of consolidated and fluid abilities, which has been observed within screening tests in groups with dementia (Brugnolo et al., 2009; Duro, Simoes, Ponciano, & Santana, 2010).

The three most central nodes in the aMCI model were AVLT Delayed (strength centrality z score = 1.44), AVLT Immediate (z = 1.36), and DSS (z = 1.19). These findings suggest that episodic memory acquisition, storage and retrieval processes, and, to a lesser extent, attention/processing speed were major influences on neurocognitive functioning in the aMCI group. Accordingly, it is hypothesised that episodic memory variables will be most central in future confirmatory network studies of patients with aMCI (hypothesis 1).

An edge between category fluency and AVLT Immediate was present in the aMCI network model, as it was in the CN network. This association (bootstrapped mean edge weight = .18, 95% CI = .11, .30) is consistent with the finding that semantic knowledge can provide scaffolding for the acquisition of episodic memories in medial temporal lobe (MTL) amnesia (Kan et al., 2009). On this basis, it is hypothesized that the semantic processes underlying category fluency performance may support the acquisition of word list memoranda in aMCI (hypothesis 4).

Overall, the characteristics of the aMCI network model suggest that the neurocognitive functioning of the aMCI group was strongly influenced by episodic memory functioning and processing speed. Additionally, while relatively distinct neurocognitive domains were evident, neurocognitive status was also shaped by the presence of memory-semantic-language and attention-speed-working memory constellations of variables.

3.3. Early Alzheimer's disease

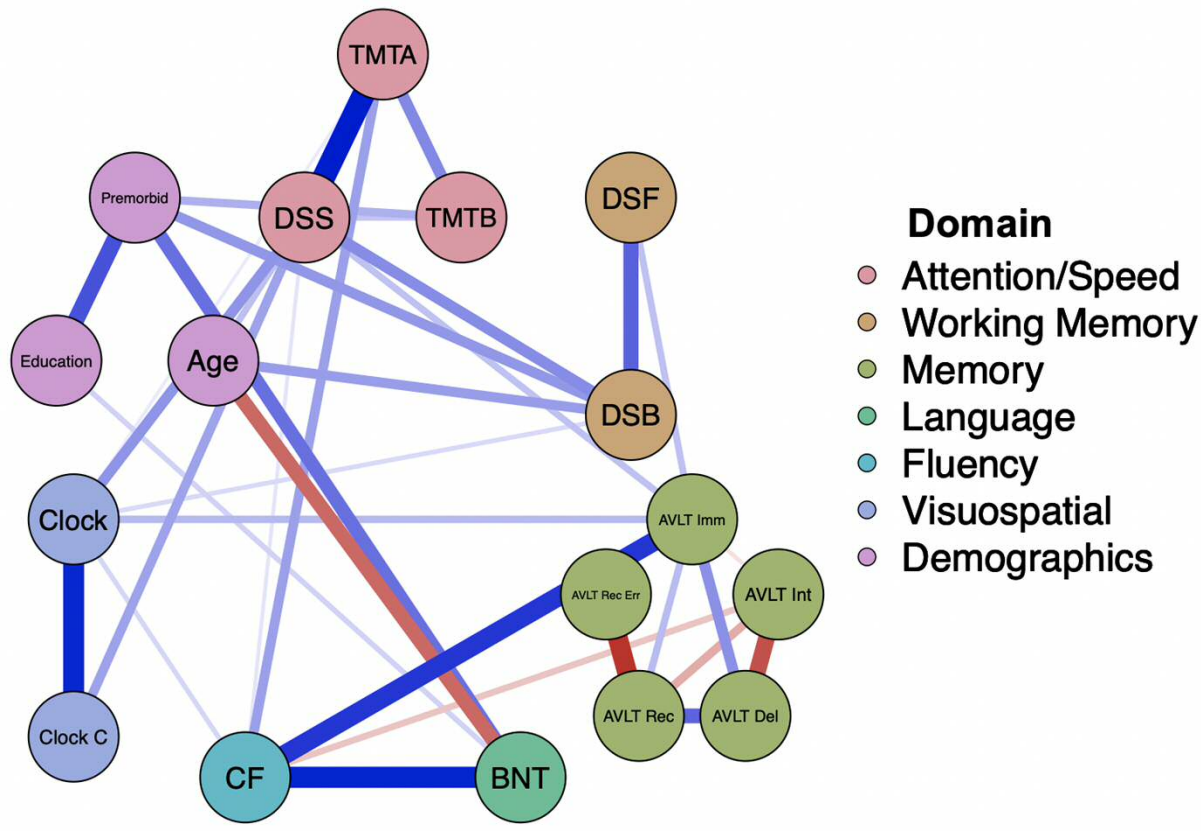
The eAD network model, shown in Fig. 3, was relatively sparse (edge density = .5), although this may be because the sample size was smaller than that of the aMCI network and EBIC is a function of N (Epskamp & Fried, 2018). The global strength of the model was not surprisingly different from the CN (NCT global strength comparison: p = .61, test statistic = .50) or the aMCI (NCT global strength comparison: p = .96, test statistic = .06) models. By contrast, the network structure of the eAD model appeared to be somewhat different from that of the CN model (NCT network structure comparison: p = .02, test statistic = .38) and the aMCI model (NCT network structure comparison: p = .02, test statistic = .38), suggesting that reorganisation occurred amongst the neurocognitive variables in the more advanced disease state.

Descriptively, the eAD model contained few weak edges between tests of different domains and a small number of larger edges linking tests of different domains (e.g., CF-BNT, bootstrapped mean edge weight = .36, 95% CI = .29, .52). These properties suggest that, at a group level, the neurocognitive functions probed by the tests did not have many slight influences on each other. Instead, they had a small number of more substantial influences on each other in this more advanced disease state.

Regarding strength centrality, DSS (strength centrality z score = 2.00), category fluency (z = 1.31), and AVLT Immediate (z = 1.01) were the three most central nodes in the model. This indicates that attention/processing speed and fluency, as well as episodic memory, were particularly strong influences on neurocognitive functioning in eAD.

Putative clusters of memory-semantic-language variables and attention-speed-working memory variables emerged in the eAD model, as was the case in the aMCI model. In the eAD model, these constellations were generally formed of stronger edges between tests, regardless of whether they probed the same or different domains. This may suggest that the neurocognitive status of the eAD group was more strongly influenced by the reorganisation of neurocognitive abilities into two main elements.

Figure 3. Neurocognitive network model for older adults with early Alzheimer's disease



DSS: Digit Symbol Substitution. TMTA: Trail Making Test Part A. TMTB: Trail Making Test Part B. DSF: Digit Span Forward. DSB: Digit Span Backward. AVLT Imm: Rey Auditory Verbal Learning Test Immediate. AVLT Imm: Rey Auditory Verbal Learning Test Intrusions. AVLT Del: Rey Auditory Verbal Learning Test Delayed. AVLT Rec: Rey Auditory Verbal Learning Test Recognition. AVLT Rec Err: Rey Auditory Verbal Learning Test Recognition Errors. BNT: Boston Naming Test. CF: category fluency. Clock: clock drawing. Clock C: clock copying. Education: education (years). Age: age (years). Premorbid: American Adult National Reading Test.

Reorganisation was also indicated within the memory-semantic-language collection of variables, especially amongst the episodic memory variables. AVLT Intrusions and recognition errors were strongly negatively associated with AVLT Delayed (bootstrapped mean edge weight = .31, 95% CI = .46, = .16) and AVLT Recognition (bootstrapped mean edge weight = .42, 95% CI = .54, = .21), respectively. By contrast, memory errors were positively associated with tests of episodic memory in

the CN model and only intrusions were negatively associated with recognition and delayed recall in the aMCI model. Subsequently, memory errors may have a more detrimental influence on episodic memory functioning and, by extension, general neurocognitive status in eAD than in aMCI and CN.

Again, a possible compensatory relationship was shown in the link between category fluency and AVLT Immediate in the eAD model. This link was present in all models, as previously discussed, but strongest for the eAD group (boot strapped mean edge weight = .36, 95% CI = .24, .49). This increase in edge strength across the eAD continuum may indicate increasing recruitment of the semantic processes underlying category fluency to support the acquisition of word list memoranda in eAD, thus offering some compensation for episodic memory impairment (hypothesis 4).

In summary, the properties of the eAD network model suggest that the neurocognitive functioning of this group was characterised by the strong influence of attention/processing speed and fluency as well as episodic memory, and a broader tendency for abilities to group into memory-semantic-language-related variables and attention-speed-working memory-related variables.

4. General discussion

Descriptions of the typical pattern of impairment in early AD (Howieson et al., 2008; Kolb, 2015; Lezak et al., 2012; Rushing et al., 2014; Weintraub et al., 2012) invoke relationships between variables. These relationships had not been explicitly statistically modelled at multiple points along the eAD continuum. Accordingly, the primary aims of the study were to estimate neurocognitive network models for CN, aMCI, and eAD participants and to explore the centrality and other network characteristics of these models. Additionally, the study aimed to generate hypotheses for future (dis)confirmatory research.

The neurocognitive network models suggested that memory variables were highly influential for all groups, particularly for the aMCI group where the immediate and delayed AVLT variables displayed the highest strength centrality. DSS, a measure of attention/processing speed, was also highly central in the network models for both clinical groups. Indeed, DSS showed the highest strength centrality in the eAD model. These network properties suggest that attention and processing speed, as well as episodic memory variables, were important influences on neurocognitive functioning in aMCI and eAD, particularly for the eAD group.

In general, non-episodic memory variables may become more central in the network models with disease progression. Notably, category fluency became more central in the network models across the eAD continuum. Category fluency was the second most central variable in the eAD network model (strength centrality z score = 1.31), suggesting that the processes sampled by this task were important influences on neurocognitive status in eAD. By contrast, category fluency variable was less central in the aMCI (z score = .54) and CN (z = .32) network models, implying it was less influential for these groups. In Tosi et al. (2020), category fluency was the most central node in the AD model. This could be because the AD participants in their study were more advanced (MMSE mean total = 20.87, 95% CI = 20.52, 21.23) than those in the present study (MMSE score of 23.4, 95% CI = 23.1, 23.7). Relationships amongst category fluency and other neurocognitive variables could plausibly change as cognitive deterioration occurs. Accordingly, it is hypothesized that category fluency becomes more central in neurocognitive network models further along the AD continuum (hypothesis 2).

While impairment on category fluency tasks is often present early in the course of AD (Adlam, Bozeat, Arnold, Watson, & Hodges, 2006; Gomez & White, 2006; Vaughan, Coen, Kenny, & Lawlor, 2018), it is unclear whether problems with executive access and/or semantic integrity underlie this. Some studies suggest that semantic knowledge appears to be relatively spared in comparison to episodic memory in the early stages of the disease (Balthazar, Cendes, & Damasceno, 2008; Linet al., 2014; Rogers & Friedman, 2008), especially regarding superordinate concepts (Giffard et al., 2001). However, other studies have suggested that significant semantic degradation is present early in AD (Adlam et al., 2006; Mårdh, Nägga, & Samuelsson, 2013). Others still have suggested that both executive access and the availability of semantically related words are damaged (Joubert et al., 2010; Weakley & Schmitter-Edgecombe, 2014). Similarly, it is not yet known which processes sampled by the test best account for its high centrality in neurocognitive network models of AD. Understanding whether semantic or executive functions best account for the centrality of category fluency in these models would help to provide a more granular view of the relationships between neurocognitive variables along the eAD continuum. It is important for future network research to address this question as centrality indices can guide interventions and offer insight on how robust networks are to deficits in particular nodes (Costantini et al., 2015).

Further reorganisation of the wider neurocognitive system was observed in the clinical groups. In the clinical network models, two putative clusters of variables emerged. One consisted of memory, semantic and language-related tasks and the other was formed of tasks probing attention, processing speed and working memory. This pattern was more pronounced in the eAD network model. Here the study echoed findings reported by Tosi et al. (2020), who noted a memory cluster and an executive cluster in their AD network model. Together, these findings indicate that neurocognitive functioning in the early clinical stages of AD are influenced by the reorganisation of neurocognitive functions into two main constellations - in contrast to the more fractionated neurocognitive domains observed in groups of CN older adults. Greater confidence can be placed in the reliability of this pattern in AD as it has been observed in two relatively large samples with differing instruments probing similar neurocognitive domains. The consistency between these two network studies represents progress in the network psychometrics of neurocognition, as data-driven exploratory methods do not necessarily generate results that generalise well to other data sets (Fried & Cramer, 2017). Further, the pattern is consistent with previous research latent variable model-based research (Brugnolo et al., 2009; Duro et al., 2010). Future (dis)confirmatory research should utilise clustering algorithms to determine whether the putative clusters that were proposed in this study meet recommended criteria for 'true' clusters (Hennig, 2015). It is hypothesized that memory-semantic-language and attention-speed-working memory clusters will characterise neurocognitive functioning, at the group level, in eAD and, to a lesser extent, in aMCI (hypothesis 3).

Finally, when the network models were inspected with reference to existing evidence, analysis of specific edges suggested a potential compensatory link between category fluency and episodic memory. After conditioning on all other nodes in the models, a link between category fluency and AVLT Immediate was present in the network models for all groups. Moreover, this link was stronger in groups further along the eAD continuum, suggesting increasing reliance on semantic processes to support the acquisition of word list information with disease progression. It is known that semantic knowledge, which is probed by category fluency tasks, can provide scaffolding

for the acquisition of episodic memoranda in MTL amnesia as a result of anoxia or encephalitis (Kan et al., 2009). While category fluency is impaired in aMCI and eAD relative to healthy controls (Weakley & Schmitter-Edgecombe, 2014; see also Table 2), and in MTL amnesiacs when the category prompt recruits episodic memory (Greenberg, Keane, Ryan, & Verfaellie, 2009), it is possible that semantic abilities compensate for impaired episodic memory acquisition in aMCI and eAD. Indeed, one study indicates that mildly impaired AD patients (CDR = .5-2.0) exhibit semantic networks that are broadly comparable to healthy controls, although they show reduced availability of semantically connected words (Weakley & Schmitter-Edgecombe, 2014). Speculatively, these semantic networks may provide a source of compensation for aMCI and eAD groups during the acquisition of word lists. Alternatively, the links between category fluency and AVLT Immediate in the network models could simply be because patients who performed well on the one test also performed well on the other and vice versa. Along these lines, the stronger relationship between these nodes in the eAD group, which displayed the most severe cognitive impairment in the study, could be because there is generally less variability in performance across tests in people with poorer overall neurocognitive functioning (Binder et al., 2009). Whether or not the semantic processes underlying category fluency performance aid the acquisition of word list memoranda in aMCI and eAD provides a hypothesis for future research to test (hypothesis 4).

4.1. Limitations

There at least two potential limitations with the methods used in this study. The first concerns variable selection, which is a prominent issue in the network psychometric literature (Fried & Cramer, 2017). If a variable that is strongly related to other variables is not included in the network model, the structure is likely to change. The exclusion of variables with restricted variance (e.g., error scores for category fluency) was necessary to ensure normality for model estimation (Epskamp & Fried, 2018) and adequate variance to produce reliable between subjects results (Hedge et al., 2018). However, this is not to say that the excluded variables are not important reflectors of neurocognitive functioning across the eAD spectrum. Indeed, some neuropsychologists have argued that error scores are highly important for understanding cognitive impairment in disease states (Milberg et al., 2009). Ultimately, it is down to the researcher to carefully decide which variables to include in a network model (Fried & Cramer, 2017).

The second possible limitation concerns the model selection and regularisation procedures employed. Simulation studies were conducted to select parameters that would likely maximise specificity, potentially at the expense of sensitivity. Pragmatically, these procedures were indicated, given the need to place confidence in the reliability of the observed edges in an exploratory study. Simulation studies also suggested that a detrimental reduction in sensitivity did not occur. Nonetheless, these methods may have led to true positive edges not being included in the models. The potential loss of sensitivity has interpretive implications as the absence of evidence for an edge in a network model is not evidence of absence of the edge in the true network. As a result, all interpretations and hypotheses were generated with caution.

The study also highlights two general limitations with the application of network psychometrics to neurocognition. First, the study estimated networks on the basis of group-level cross-sectional data. It is problematic to make inferences about the neurocognitive functioning of individual patients from group level data, as group

averages do not necessarily pertain to individual patients. Second, it is harder to make causal inferences from network models based on cross-sectional, as opposed to longitudinal, data. These limitations also apply to practically all latent-variable models of neurocognition in the neuropsychological literature (Delis et al., 2003). To address these limitations, future network research could utilise within-subject longitudinal data to investigate individual differences as well as group patterns in neurocognitive functioning. Such data might be generated through experience sampling methodology, which has recently been validated for investigating relationships between processing speed and affect in a non-clinical sample (Verhagen et al., 2019). A further benefit of using longitudinal data would be the opportunity to statistically model the temporal dynamics of neurocognitive functioning, which are also invoked in descriptions of the typical patterns of impairment in eAD worsening over time.

5. Conclusion

In summary, the relationships amongst neurocognitive variables may change in a way that is hinted at but not clearly stated in the description of the typical profile of AD onset. While memory is a predominant influence on neurocognitive functioning in aMCI and eAD (and indeed in CN older adults), there was evidence of wider reorganisation of the neurocognitive system in these disease states. Namely, attention/processing speed and category fluency became more influential in eAD, and two putative clusters (memory-semantic-language and attention-speed-working memory) of neurocognitive variables emerged in the aMCI and eAD groups. This pattern was more pronounced in the more severely impaired eAD group. The findings were broadly consistent with those of Tosi et al. (2020). By including an intermediate group (aMCI) between CN and eAD, this study provided a granular view of the relationships between neurocognitive variables at three points along the eAD continuum. Additionally, it generated hypotheses for future (dis)confirmatory research.

6. Hypotheses

Four hypotheses were generated:

1. Episodic memory variables will be most central in confirmatory neurocognitive network models of aMCI.
2. Category fluency becomes increasingly central (i.e., influential) in neurocognitive network models of groups with more severe AD.
3. Memory-semantic-language and attention-speed-working memory clusters will emerge in confirmatory neurocognitive network models for aMCI, eAD and AD groups. They will be more pronounced for groups with more severe AD.
4. The semantic networks underlying category fluency performance support the acquisition of word list memoranda in aMCI and eAD.

7. Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La

Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

8. Declaration of interest

Nothing to declare.

9. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2021.01.002.

References

Adlam, A. L. R., Bozeat, S., Arnold, R., Watson, P., & Hodges, J. R. (2006). Semantic knowledge in mild cognitive impairment and mild Alzheimer's disease. *Cortex*, *42*(5), 675–684. https://doi.org/10.1016/S0010-9452(08)70404-0

Balthazar, M. L. F., Cendes, F., & Damasceno, B. P. (2008). Semantic error patterns on the Boston Naming Test in normal aging, amnestic mild cognitive impairment, and mild Alzheimer's disease: Is there semantic disruption? In *Neuropsychology* (Vol. 22, Issue 6, pp. 703–709). American Psychological Association. https://doi.org/10.1037/a0012919

Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: "abnormal" Neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, *24*(1), 31–46. https://doi.org/10.1093/arclin/acn001

Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the national adult reading test. *Clinical Neuropsychologist*, *3*(2), 129–136. https://doi.org/10.1080/13854048908403285

Bowden, S., Cook, M., Bardenhagen, F., Shores, E., & Carstairs, J. (2004). Measurement invariance of core cognitive abilities in heterogenous neurological and community samples. *Intelligence*, *32*. https://doi.org/10.1016/j.intell.2004.05.002

Brugnolo, A., Nobili, F., Barbieri, M. P., Dessi, B., Ferro, A., Girtler, N., Palummeri, E., Partinico, D., Raiteri, U., Regesta, G., Servetto, G., Tanganelli, P., Uva, V., Mazzei, D., Donadio, S., De Carli, F., Colazzo, G., Serrati, C., & Rodriguez, G. (2009). The factorial structure of the mini mental state examination (MMSE) in Alzheimer's disease. *Archives of Gerontology and Geriatrics*, *49*(1), 180–185. https://doi.org/https://doi.org/10.1016/j.archger.2008.07.005

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

Diaz-Asper, C. M., Schretlen, D. J., & Pearlson, G. D. (2004). How well does IQ predict

neuropsychological test performance in normal adults? *Journal of the International Neuropsychological Society*, *10*(1), 82–90. https://doi.org/DOI: 10.1017/S1355617704101100

Duro, D., Simões, M. R., Ponciano, E., & Santana, I. (2010). Validation studies of the Portuguese experimental version of the Montreal Cognitive sAssessment (MoCA): confirmatory factor analysis. *Journal of Neurology*, *257*(5), 728–734. https://doi.org/10.1007/s00415-009-5399-5

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., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(May 2014). https://doi.org/10.18637/jss.v048.i04

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*(4), 617–634. 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), 1–13. https://doi.org/10.1371/journal.pone.0179891

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian Graphical Model in Cross-Sectional and Time-Series Data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Fried, E. I., & Cramer, A. O. J. (2017). Moving Forward: Challenges and Directions for Psychopathological Network Theory and Methodology. *Perspectives on Psychological Science*, *12*(6), 999–1020. https://doi.org/10.1177/1745691617705892

Giffard, B., Desgranges, B., Nore-Mary, F., Lalevée, C., De La Sayette, V., Pasquier, F., & Eustache, F. (2001). The nature of semantic memory deficits in Alzheimer's disease: New insights from hyperpriming effects. *Brain*, *124*(8), 1522–1532. https://doi.org/10.1093/brain/124.8.1522

Gomez, R. G., & White, D. A. (2006). Using verbal fluency to detect very mild dementia of the Alzheimer type. *Archives of Clinical Neuropsychology*, *21*(8), 771–775. https://doi.org/10.1016/j.acn.2006.06.012

Goodglass, H., & Kaplan, E. (1983). The assessment of aphasia and related disorders. Philadelphia: Lea and Febiger.

Greenberg, D. L., Keane, M. M., Ryan, L., & Verfaellie, M. (2009). Impaired category fluency in medial temporal lobe amnesia: The role of episodic memory. *Journal of Neuroscience*, *29*(35), 10900–10908. https://doi.org/10.1523/JNEUROSCI.1202-09.2009

Greenberg, D. L., & Verfaellie, M. (2010). Interdependence of episodic and semantic memory: Evidence from neuropsychology. *Journal of the International Neuropsychological Society*, *16*(5), 748–753. https://doi.org/10.1017/S1355617710000676

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hennig, C. (2015). What are the true clusters? *Pattern Recognition Letters*, *64*, 53–62. https://doi.org/10.1016/j.patrec.2015.04.009

Howieson, D. B., Carlson, N. E., Moore, M. M., Wasserman, D., Abendroth, C. D., Payne-Murphy, J., & Kaye, J. A. (2008). Trajectory of mild cognitive impairment onset. *Journal of the International Neuropsychological Society*, *14*(2), 192–198. https://doi.org/DOI: 10.1017/S1355617708080375

Joubert, S., Brambati, S. M., Ansado, J., Barbeau, E. J., Felician, O., Didic, M., Lacombe, J., Goldstein, R., Chayer, C., & Kergoat, M. J. (2010). The cognitive and neural expression of semantic memory impairment in mild cognitive impairment and early Alzheimer's disease. *Neuropsychologia*, *48*(4), 978–988. https://doi.org/10.1016/j.neuropsychologia.2009.11.019

Kan, I. P., Alexander, M. P., & Verfaellie, M. (2009). Contribution of prior semantic knowledge to new episodic learning in amnesia. *Journal of Cognitive Neuroscience*, *21*(5), 938–944. https://doi.org/10.1162/jocn.2009.21066

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). The Boston Naming Test. Philadelphia: Lea and Febiger.

Kolb, B. (2015). Fundamentals of human neuropsychology. In I. Q. Whishaw (Ed.), *Human neuropsychology* (Seventh ed). New York : Worth Publishers.

Lauritzen, S. (1996). *Graphical models*. Clarendon press.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). Neuropsychological assessment, 5th ed. In *Neuropsychological assessment, 5th ed.* Oxford University Press.

Lin, C.-Y., Chen, T.-B., Lin, K.-N., Yeh, Y.-C., Chen, W.-T., Wang, K.-S., & Wang, P.-N. (2014). Confrontation naming errors in Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *37*(1–2), 86–94. https://doi.org/10.1159/000354359

Liu, H., Lafferty, J., & Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, *10*, 2295–2328. https://doi.org/10.1184/r1/6610712

Mårdh, S., Nägga, K., & Samuelsson, S. (2013). A longitudinal study of semantic memory impairment in patients with Alzheimer's disease. *Cortex*, *49*(2), 528–533. https://doi.org/10.1016/j.cortex.2012.02.004

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer{\textquoteright}s disease. *Neurology*, *34*(7), 939. https://doi.org/10.1212/WNL.34.7.939

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Measurement*, *58*(4), 525–543.

Milberg, W. P., Hebben, N., & Kaplan, E. (2009). The Boston Process Approach to neuropsychological assessment. In *Neuropsychological assessment of neuropsychiatric and neuromedical disorders, 3rd ed.* (pp. 42–65). Oxford University Press.

Mohn, C., Sundet, K., & Rund, B. R. (2014). The relationship between IQ and performance on the MATRICS consensus cognitive battery. *Schizophrenia Research: Cognition*, *1*(2), 96–100. https://doi.org/10.1016/j.scog.2014.06.003

Moroney, J. T., Bagiella, E., Desmond, D. W., Hachinski, V. C., Mölsä, P. K., Gustafson, L., Brun, A., Fischer, P., Erkinjuntti, T., Rosen, W., Paik, M. C., & Tatemichi, T. K. (1997). Meta-analysis of the Hachinski Ischemic Score in pathologically verified dementias. *Neurology*, *49*(4), 1096–1105. https://doi.org/10.1212/wnl.49.4.1096

Morris, J. C. (1993). The Clinical Dementia Rating (CDR). *Neurology*, *43*(11), 2412--2412-a. https://doi.org/10.1212/WNL.43.11.2412-a

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G.,Fillenbaum, G.,...Clark, C. (1989). The Consortium to establisha eegistry for Alzheimer's Disease (CERAD). Part I. Clinical andneuropsychological assessment of Alzheimer's disease.Neurology, 30(9), 1159e1165. https://doi.org/10.1212/wnl.39.9.1159

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251. https://doi.org/10.1016/j.socnet.2010.03.006

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jack, C. R., Jagust, W. J., Shaw, L. M., Toga, A. W., Trojanowski, J. Q., & Weiner, M. W. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology*, *74*(3), 201–209. https://doi.org/10.1212/WNL.0b013e3181cb3e25

Reitan, R. (1958). Validity of the trail making test as an indicator of organic brain damage. Perceptual and Motor Skills, 8(3), 271e276. https://doi.org/10.2466/pms.1958.8.3.271

Rey, A. (1964). L'examen clinique en psychologie. Paris: Presses universitaires de France.

Rogers, S. L., & Friedman, R. B. (2008). The underlying mechanisms of semantic memory loss in Alzheimer's disease and semantic dementia. *Neuropsychologia*, *46*(1), 12–21. https://doi.org/10.1016/j.neuropsychologia.2007.08.010

Rosen, W. G., Terry, R. D., Fuld, P. A., Katzman, R., & Peck, A. (1980). Pathological verification of ischemic score in differentiation of dementias. *Annals of Neurology*, *7*(5), 486–488. https://doi.org/10.1002/ana.410070516

Rushing, N. C., Sachs-Ericsson, N., & Steffens, D. C. (2014). Neuropsychological indicators of preclinical Alzheimer's disease among depressed older adults. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *21*(1), 99–128. https://doi.org/10.1080/13825585.2013.795514

Strauss, M. E., & Smith, G. T. (2009). Construct Validity: Advances in Theory and Methodology. *Annual Review of Clinical Psychology*, *5*(1), 1–25. https://doi.org/10.1146/annurev.clinpsy.032408.153639

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. http://www.jstor.org/stable/2346178

van Bork, R., van Borkulo, C. D., Waldorp, L. J., Cramer, A. O. J., & Borsboom, D. (2018). Network Models for Clinical Psychology. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (Issue May, pp. 1–35). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119170174.epcn518

van Borkulo, C. D., Boschloo, L., Kossakowski, J. J., Tio, P., Schoevers, R. A., Borsboom, D., & Waldorp, L. J. (2017). Comparing network structures on three aspects: A permutation test. *Manuscript Submitted*, *March*, 34. https://doi.org/10.13140/RG.2.2.29455.38569

Vaughan, R. M., Coen, R. F., Kenny, R., & Lawlor, B. A. (2018). Semantic and Phonemic Verbal Fluency Discrepancy in Mild Cognitive Impairment: Potential Predictor of Progression to Alzheimer's Disease. *Journal of the American Geriatrics Society*, *66*(4), 755–759. https://doi.org/10.1111/jgs.15294

Weakley, A., & Schmitter-Edgecombe, M. (2014). Analysis of verbal fluency ability in Alzheimer's disease: The Role of clustering, switching and semantic proximities. *Archives of Clinical Neuropsychology*, *29*(3), 256–268. https://doi.org/10.1093/arclin/acu010

Wechsler, D. (1981). Wechsler Adult Intelligence Scale - Revised. San Antonio: Psychological Corporation

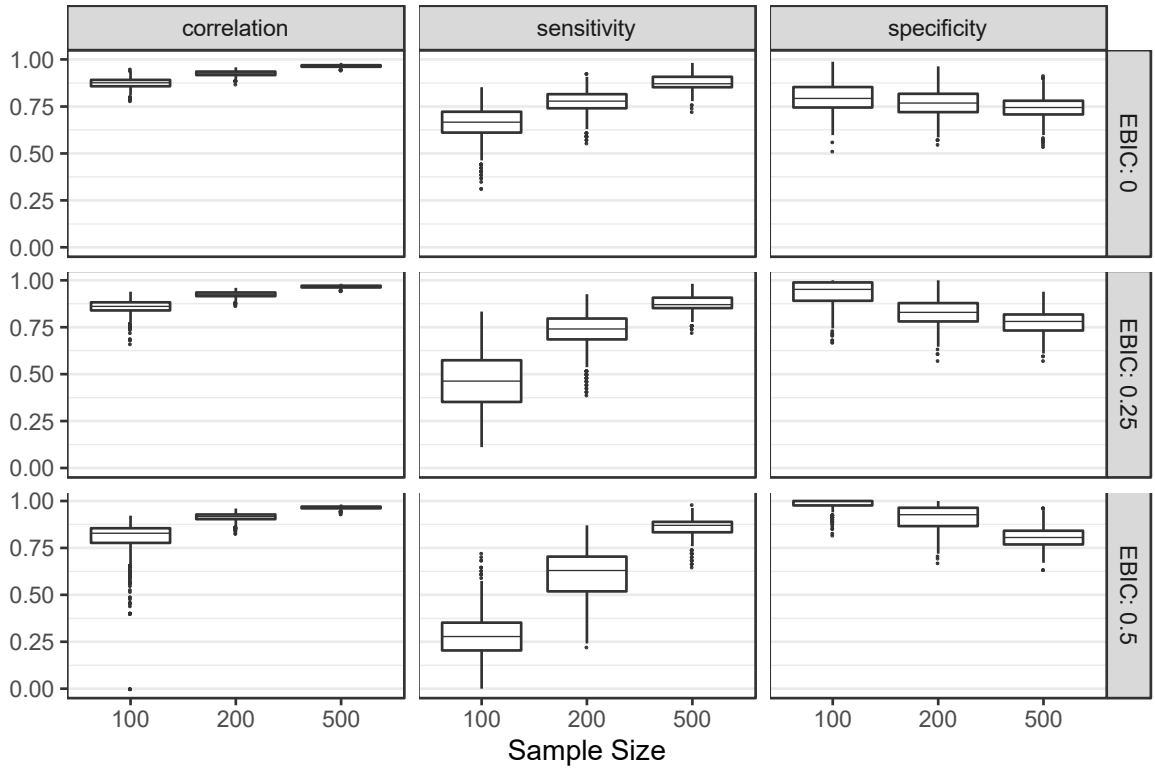
Wechsler, D. (1987). Wechsler Memory Scale-Revised. *Psychological Corporation*. http://ci.nii.ac.jp/naid/10008745172/en/

Weintraub, S., Wicklund, A. H., & Salmon, D. P. (2012). The neuropsychological profile of Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, *2*(4), a006171–a006171. https://doi.org/10.1101/cshperspect.a006171

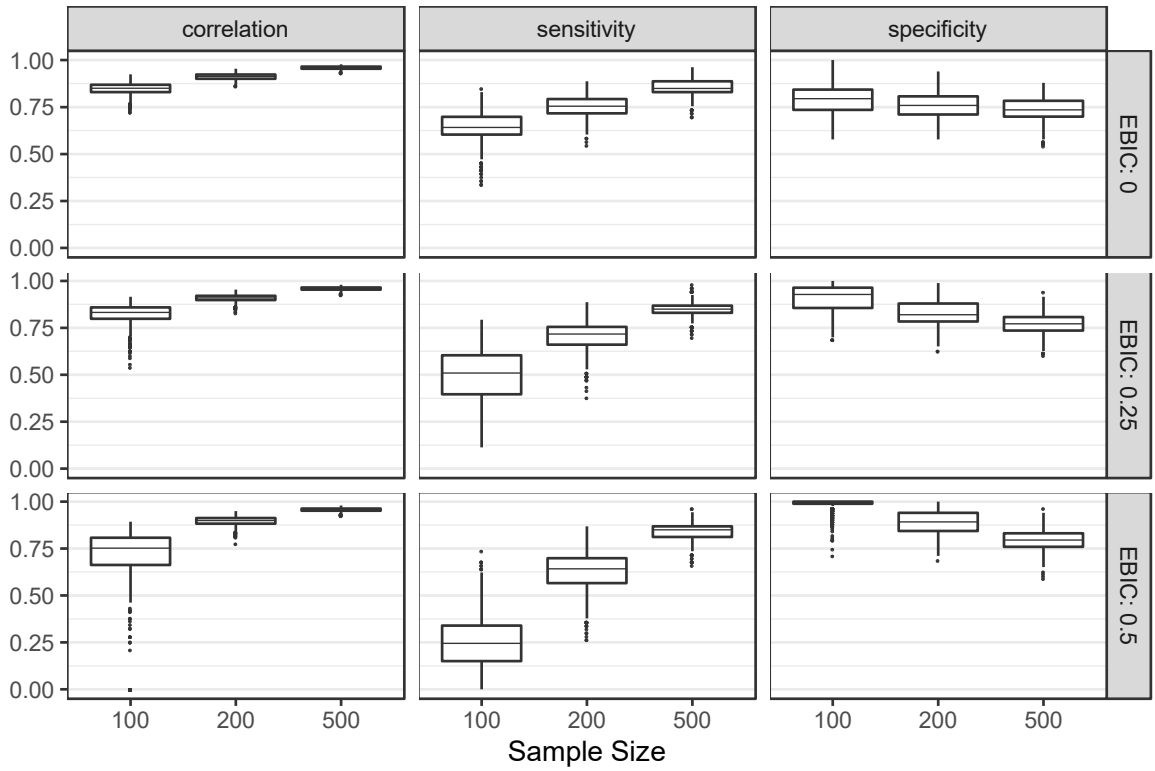
Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1983). Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatric Research*, *17*(1), 37–49. https://doi.org/10.1016/0022-3956(82)90033-4

Appendix A: simulation studies of model selection and regularisation parameters

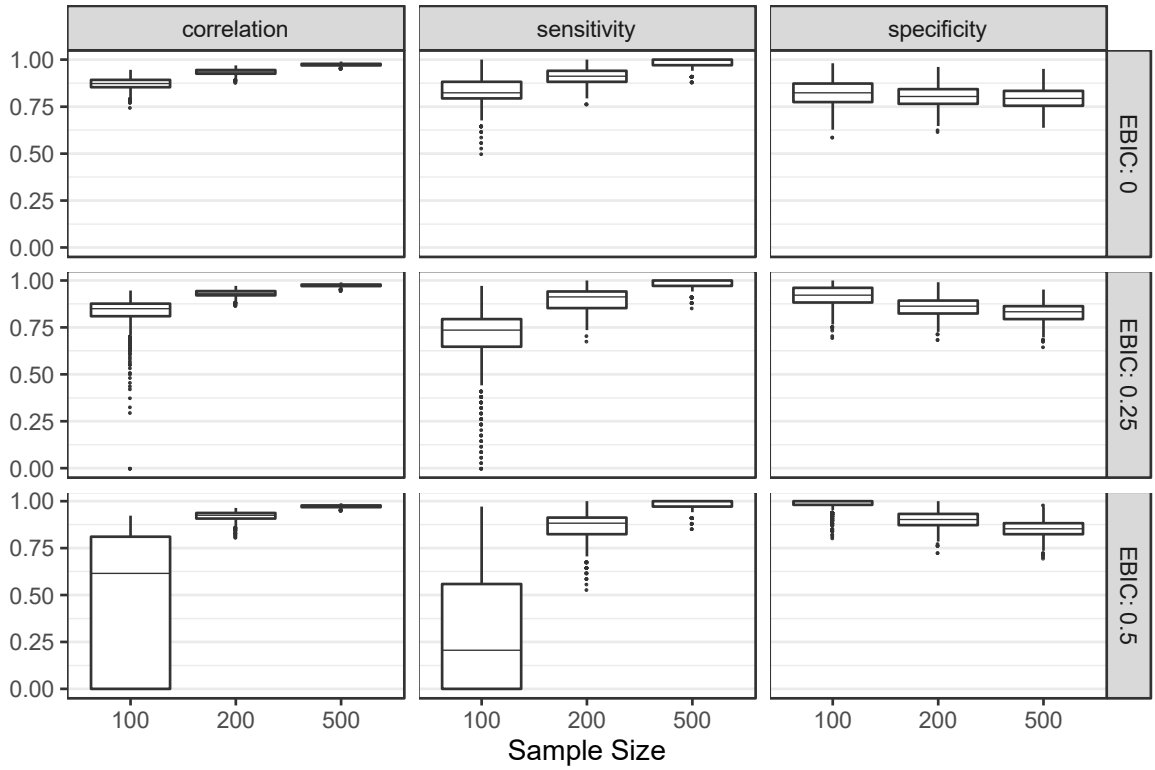
CN network model simulation study



aMCI network model simulation study

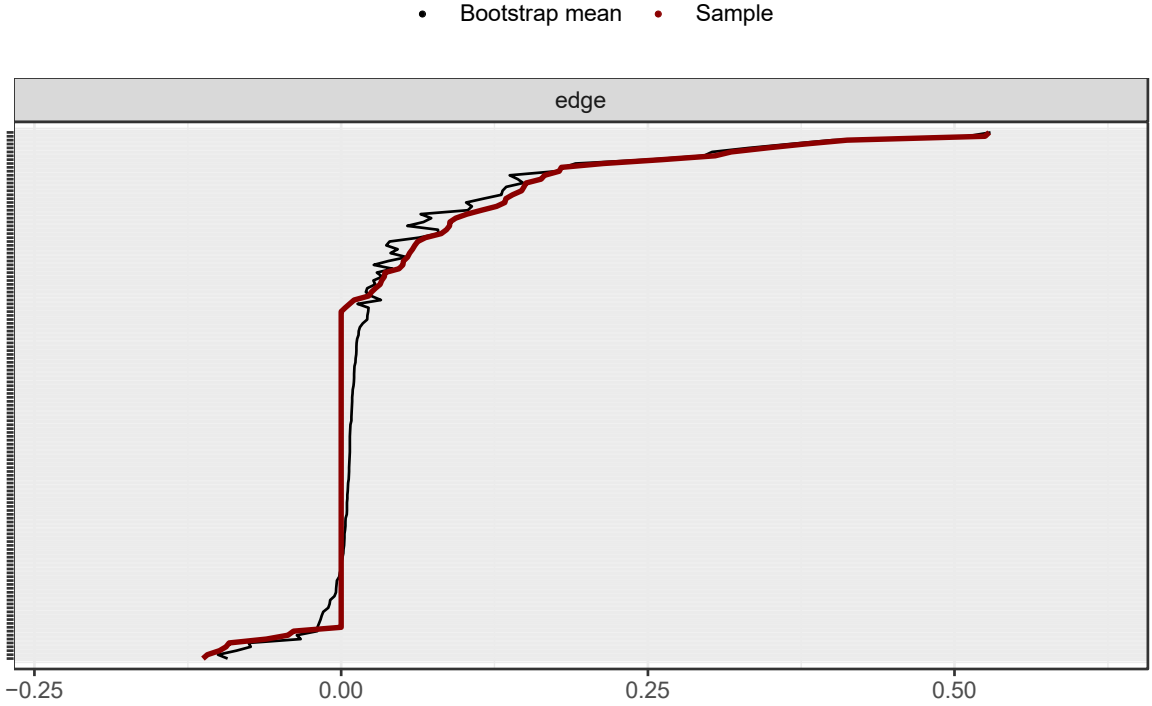


eAD network model simulation study

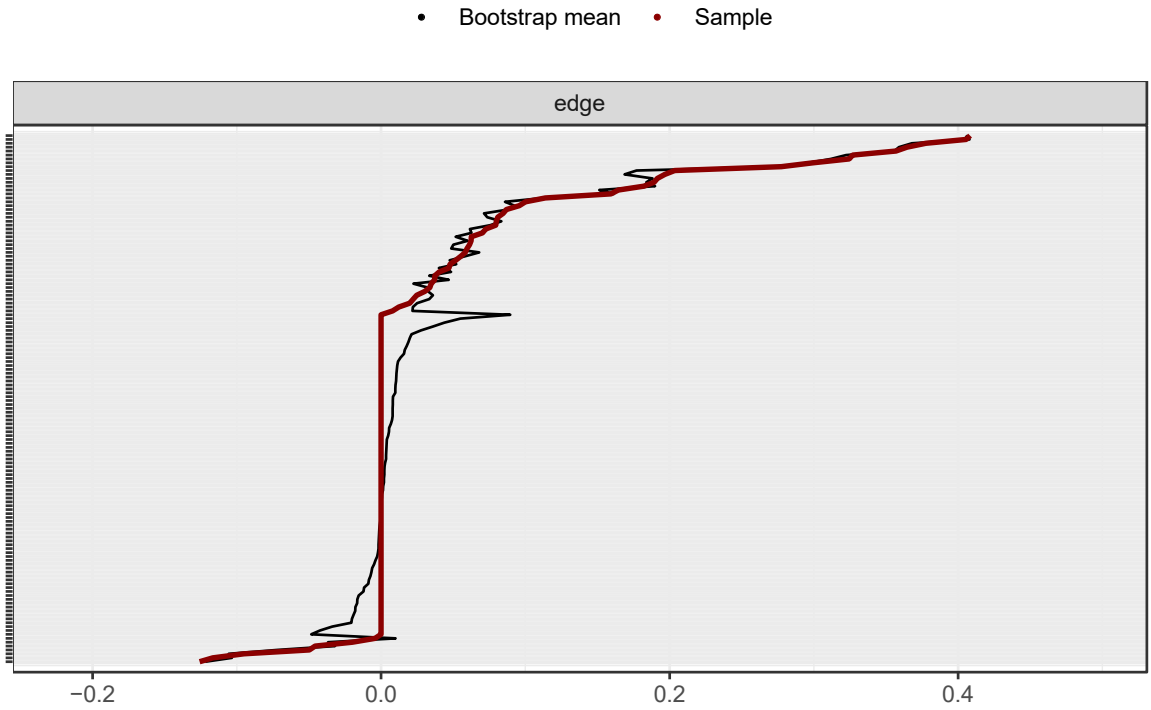


Appendix B: edge weight bootstraps

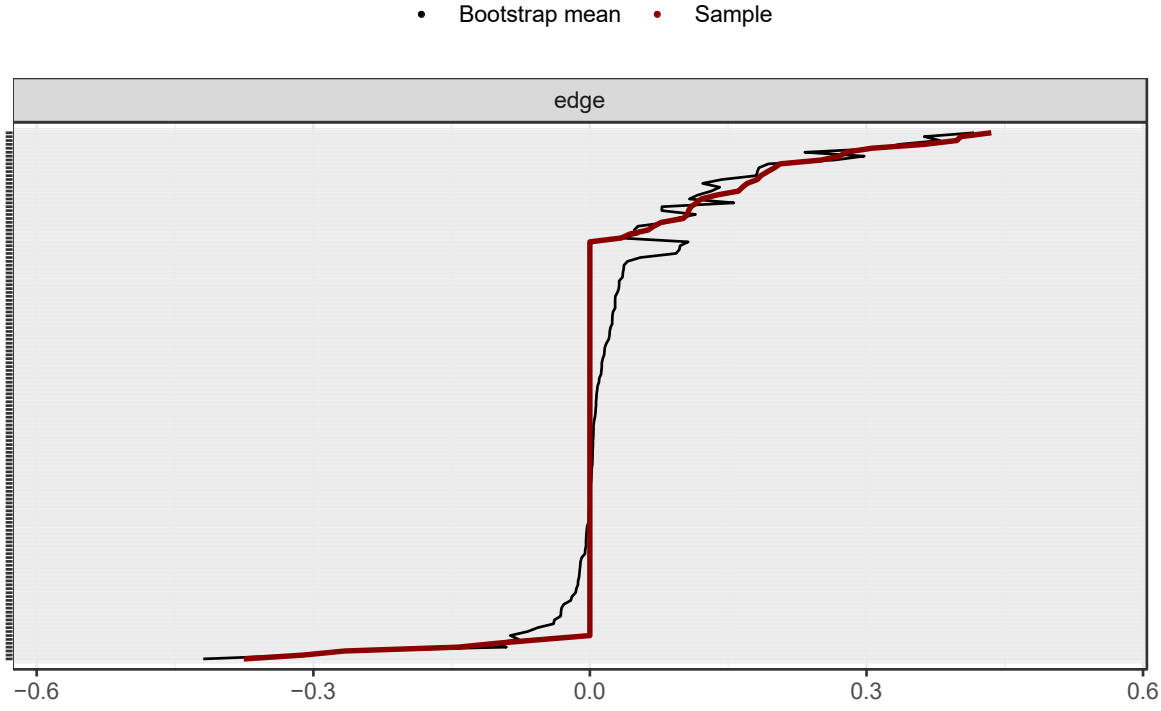
Confidence intervals around regularized edge weights in the CN network model



Confidence intervals around regularized edge weights in the aMCI network model

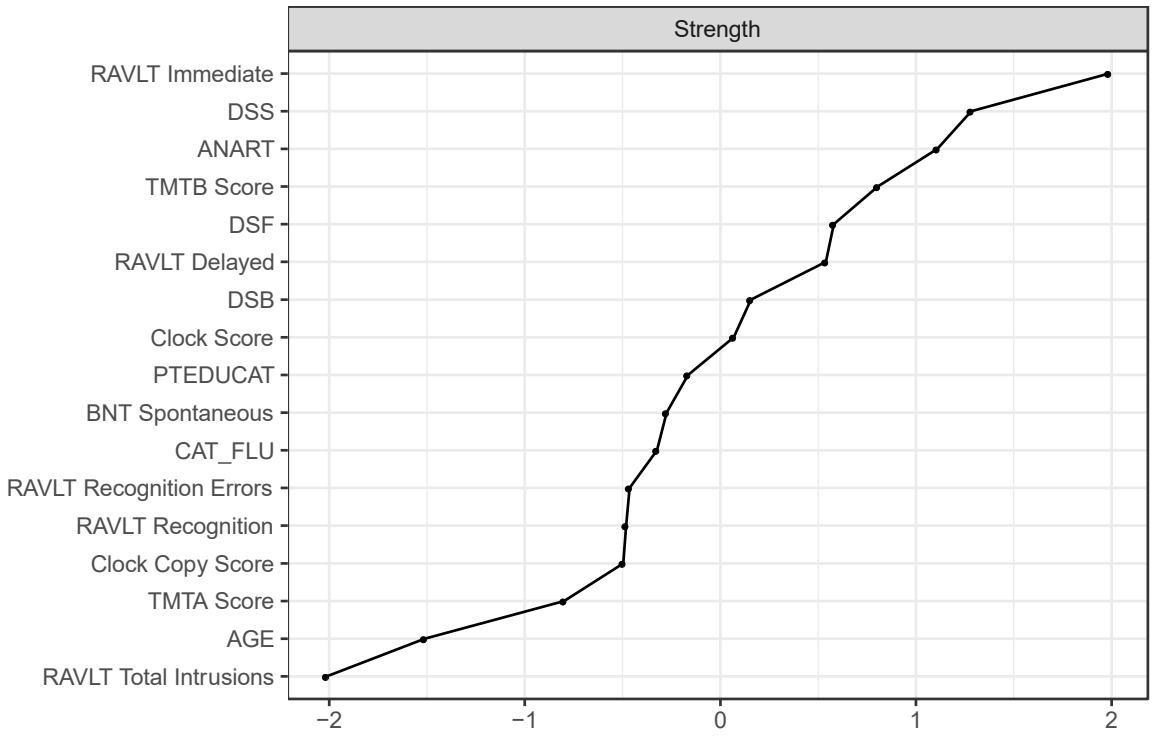


Confidence intervals around regularized edge weights in the eAD network model

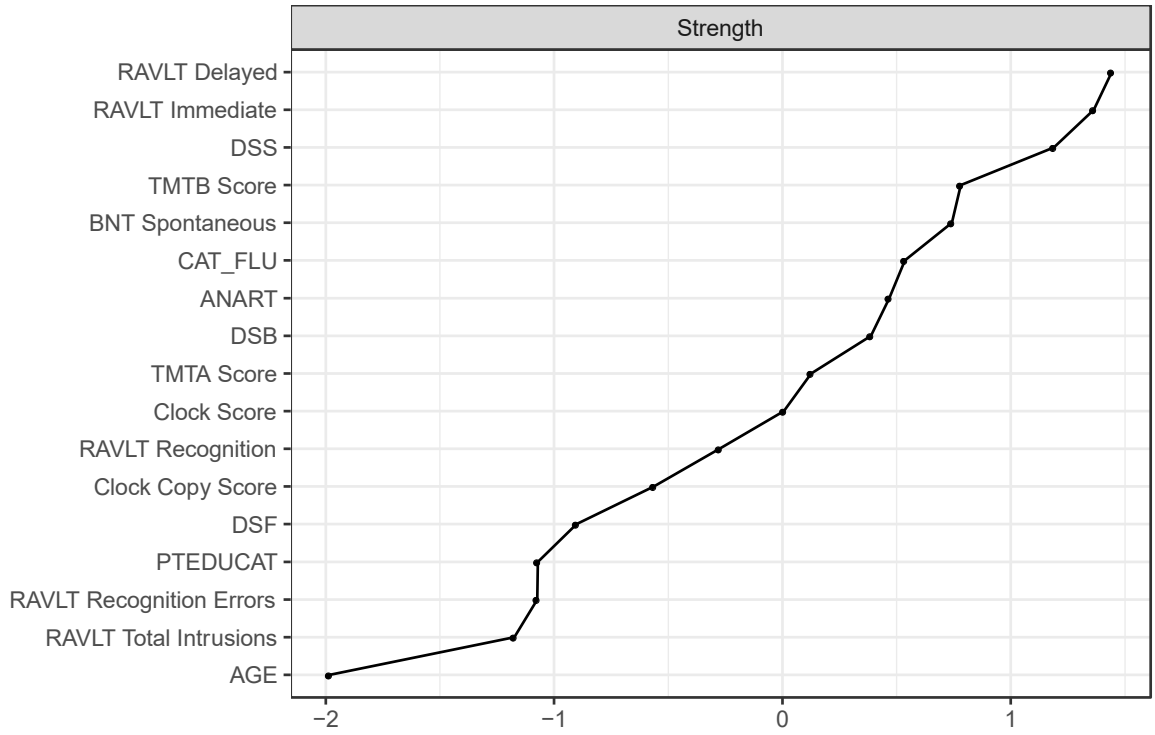


Appendix C: centrality estimates

CN network model centrality estimates



aMCI network model centrality estimates



eAD network model centrality estimates