Received: 7 April 2026 | Accepted: 12 August 2026

DOI: 10.1111/jnp.70076

the british psychological society
promoting excellence in psychology

REVIEW ARTICLE

Estimating and interpreting psychometric networks in neuropsychology: A tutorial with conceptual extensions

Giorgia Tosi | **Daniele Romano**

Psychology Department, University of Milan-Bicocca, Milan, Italy

Correspondence
Daniele Romano and Giorgia Tosi, Psychology Department, University of Milan-Bicocca, Milan, Italy.
Email: daniele.romano@unimib.it and giorgia.tosi@unimib.it

Abstract

Network psychometrics has emerged as a promising framework for modelling the structure of neuropsychological data, offering a complementary perspective to traditional latent variable approaches or single test approaches. In neuropsychology, test performance is rarely interpreted in isolation, but rather as reflecting interacting cognitive processes. Network models provide a formal framework to represent these interdependencies by estimating conditional associations among observed variables. The present article provides a comprehensive tutorial on the estimation and interpretation of psychometric networks in neuropsychology. First, we offer a step-by-step guide to estimating a cross-sectional Gaussian Graphical Model (GGM), including data preparation, model selection, regularization, visualization and stability assessment, using a neuropsychological dataset from patients with Alzheimer's disease. Particular emphasis is placed on best practices for ensuring reproducibility and robustness, including the use of bootstrap procedures and reporting standards. Second, we outline key methodological extensions, including Exploratory Graph Analysis for dimensionality assessment, joint graphical models and Network Comparison Test for group comparisons, mixed graphical models for heterogeneous data and longitudinal network approaches. These methods are presented conceptually to illustrate how network analysis can be extended beyond basic cross-sectional applications. Finally, we discuss the strengths and limitations of network psychometrics in neuropsychology. We emphasize that network models represent statistical associations among observed test scores rather than direct representations of cognitive architecture and that their interpretation requires theoretical caution. When applied with methodological rigour and

conceptual restraint, network analysis provides a flexible and integrative framework for investigating the structure of cognitive systems.

KEYWORDS

cognitive networks, cognitive system, exploratory graph analysis, Gaussian graphical models, network psychometrics, neuropsychological assessment

FROM NEUROPSYCHOLOGICAL TEST SCORES TO COGNITIVE SYSTEMS: WHY NETWORK PSYCHOMETRICS?

Neuropsychological assessment has always operated under a productive tension. On the one hand, individual tests are designed to isolate specific cognitive functions, for example episodic memory, working memory, attention, language, or executive control. This domain-based organization is clinically useful: it guides test selection, supports diagnostic reasoning and allows clinicians to describe patterns of impairment in an interpretable way. On the other hand, performance on any given task rarely reflects a single isolated function. Memory performance depends on attentional efficiency and executive strategies; language comprehension can influence verbal learning; visuospatial and executive demands may jointly affect constructional tasks; and demographic variables such as age and education modulate performance across multiple domains. Thus, while neuropsychological diagnostics often uses modular categories, clinical interpretation frequently requires integrating information across tests and domains. Network psychometrics has emerged over the past decade as one such framework (Borsboom et al., 2021). Rather than assuming that correlations among observed variables arise solely from latent common causes, network models represent variables as mutually interacting components within a structured system. In this approach, observed variables—such as neuropsychological test scores—are conceptualized as nodes in a network and edges represent statistical associations between them, typically conditional dependencies (Ferguson, 2022; Tosi et al., 2020). Recent methodological work has clarified that network and latent variable models are not necessarily incompatible, but may describe different aspects of the same data-generating process (Briganti et al., 2024).

Network psychometrics is useful precisely because it provides a statistical framework for representing this dual nature of neuropsychological data. It does not require abandoning cognitive domains or modular interpretations. Rather, it allows researchers to examine how domain-specific measures are conditionally related within a broader cognitive system. From this perspective, cognitive domains can still be used as clinically meaningful organizing concepts, but their relationships, overlaps and possible reorganization can be explicitly modelled. Network models, therefore, offer a way to move beyond interpreting tests as either isolated indicators or simple manifestations of latent domains and instead represent cognition as a complex, interconnected system.

The relevance of network approaches for neuropsychology is not merely theoretical. Empirical applications have begun to demonstrate how cognitive systems can be characterized structurally. For example, Tosi et al. (2020) applied Gaussian Graphical Models to neuropsychological batteries in healthy aging, Alzheimer's disease and vascular encephalopathy, revealing that cognitive impairment may involve not only reduced performance but also reorganization of inter-test relationships. On the contrary, healthy individuals displayed more segregated structures consistent with domain specificity, whereas patient groups exhibited denser and altered connectivity patterns, suggesting systemic rearrangement. Similarly, Zoccolotti and colleagues (Zoccolotti et al., 2021) modelled reading, spelling and mathematical skills in children, identifying both domain-specific clustering and cross-domain associations consistent with shared acquisition mechanisms. Other neuropsychological network studies have reported

different patterns of global connectivity across clinical groups, while still converging on the broader idea that neurological disease may alter the architecture of cognitive relationships rather than only lowering test scores (e.g. De Marco et al., 2024; Ferguson & Foley, 2024; Scharfenberg et al., 2025; Wright et al., 2021). These findings converge on an important insight: neuropsychological profiles are structured systems. Pathology may alter the architecture of relationships among cognitive components, not merely depress individual node values. A methodological framework capable of modelling such architecture is therefore of direct clinical and theoretical relevance.

Network analysis can contribute to several neuropsychological questions. It can be used to examine whether cognitive domains form relatively segregated or integrated systems, whether clinical groups differ in the organization of cognitive relationships, whether impairment is associated with systemic reorganization, and whether intervention changes the structure of associations among cognitive or clinical variables. Thus, network models are not only descriptive tools for visualizing correlations among test scores, but also a framework for formulating and testing hypotheses about cognitive organization, clinical heterogeneity and change. As emphasized in the broader network literature, particularly in clinical psychology, network models risk becoming purely descriptive when applied without a clear theoretical or clinical question (Fried, 2020). In this perspective, network analysis is most useful when it is used to formalize explicit questions about cognitive organization. For example, researchers may ask whether cognitive impairment reflects a reorganization of relationships among functions, whether clinical groups differ in the architecture of cognitive associations, whether observed test scores cluster into theoretically meaningful domains, or whether rehabilitation modifies the structure of cognitive or clinical variables. Thus, the methodological choice should follow the neuropsychological question, rather than the reverse.

The aim of the present tutorial is twofold. First, we provide a step-by-step guide to estimating and interpreting a basic cross-sectional psychometric network using contemporary best practices, including regularized Gaussian Graphical Models and stability analysis. This section is designed to bring the reader, methodically and cautiously, to the point of constructing a methodologically sound first network. Second, we introduce a set of advanced extensions—conceptual rather than procedural—that illustrate how network methods can be expanded to address dimensionality, group comparisons, longitudinal dynamics, mixed data types and Bayesian inference.

HOW TO CREATE A NEUROPSYCHOLOGICAL NETWORK: A STEP-BY-STEP GUIDE

We will present the general workflow for conducting a network analysis and provide a step-by-step tutorial (see Supplemental Materials) in the statistical software R, version 4.4.1 (2024-06-14; R Core Team, 2021). A glossary of all relevant concepts for understanding network psychometrics is also provided (see Supplemental Materials). Box 1 shows the network analysis reporting standards based on Burger et al. (2023).

The example runs on cross-sectional data from a recent study on cognitive reorganization in dementia (Tosi et al., 2024). The sample consists of $N = 241$ patients with probable Alzheimer's Disease (148 females; mean age, 72.2 ± 8.03 years; mean years of education, 7.69 ± 4.50). The dataset (supplemental file = *neuropsychological_example.csv*) includes a neuropsychological assessment comprising the following tests: Minimental State Examination (MMSE; Folstein et al., 1975); Frontal Assessment Battery (FAB; Dubois et al., 2008); semantic fluency test (SemF; Capasso & Miceli, 2001); phonemic fluency test (PhonF; Caltagirone et al., 1995); Boston naming test (BNT; Kaplan et al., 2001); digit span forward and backward (DSf and DSb; Monaco et al., 2013); Rey Auditory Verbal Learning Test immediate and delayed recall (RVLTi and RVLTd; Caltagirone et al., 1995); clock drawing test (CDT; Caffarra et al., 2011; Freedman et al., 1994); and incomplete letters subtest of the visual object and space perception battery (VSil; Warrington & James, 1991). See Tosi et al. (2024) for a complete description of the data. A fully commented script is available in Supplemental Materials (*tutorial script.Rmd*). When the text refers to precise locations in the code, it

BOX 1 Reporting standards. The box shows the reporting standards based on Burger et al. (2023)

Reporting standards (Burger et al., 2023)

Model and variable selection. In the measure section, describe variables by their distributions, report any violations of assumptions, and note any data transformations, imputation or removal of missing data (Burger et al., 2023). Also, indicate the final number of variables included in the network analysis and the selected model (i.e. GGM).

Network structure estimation. In the analysis section, specify the estimation method used (e.g. EBICglasso), any additional specifications (i.e. the chosen threshold, regularization parameters, and correlation method), and the default arguments, because defaults in software packages can change. Finally, there is a recommendation for reporting the statistical software and packages used, including their versions.

Visualization and results. In the results section, include the inverse covariance matrix, along with the zero-order correlation matrix. When a network plot is included in the article, report what the edges represent (i.e. partial correlations), the size of the smallest and largest edges in the network, the function used to create the plot (e.g. *qgraph*), the layout setting, and the colour code.

Network description and metrics. In the results section, report basic descriptives of the network structure (i.e. density and mean weight). In addition to visualizing centrality indices, include a table showing the raw centrality scores, as exact parameter values can often not be inferred from centrality plots.

Network stability. In the analysis section, report accuracy analysis both for edges and centrality indices, specifying the bootstrapping methods used and the number of resamplings computed. In the results section, report at least the edge-stability plot and the case-drop bootstrap plot for the reported centrality indices. When possible, also include the relative tables. Also, report the correlation-stability (CS) coefficient.

will be indicated as follows: the *neuropsychological_example.csv* dataset can be loaded into the R environment (R lines 38–46), and descriptive statistics can be explored (R lines 48–49).

R script and Supplemental Materials are available on the Open Science Platform at the following link: https://osf.io/73y5r/overview?view_only=35ff46d85a904506bf98b364b833f68f.

A priori sample size determination and power considerations

Determining the appropriate sample size for network analysis is complex and debated in the literature, as it depends on factors such as the number of nodes (variables), the density of the true network and the strength of the edges (Constantin, 2019; Constantin et al., 2026). The required sample size primarily depends on the complexity of the model being estimated (Ávalos-Tejeda & Calderón, 2025), and recommendations vary depending on the study design. For cross-sectional networks, standard power analysis is difficult because the 'true' network structure is rarely known a priori. However, simulation studies suggest that for networks of moderate size (<20 nodes), sample sizes of 250–350 are typically required to achieve acceptable sensitivity and specificity (Constantin, 2019). A common heuristic is that the sample size should be at least three times the number of parameters being estimated, though Monte Carlo simulations are preferred for precise planning (Constantin et al., 2026).

Although precise recommendations are lacking, and given the relatively small sample sizes that typically characterize psychological research, it is always recommended to determine the stability and

robustness of the estimated network (Epskamp, Borsboom, & Fried, 2018; Hevey, 2018). Regardless of sample size, data quality (i.e. reliability of the variables) and the underlying true structure of the network significantly impact estimation stability. Including variables with high measurement error will invariably lead to unstable edge estimates. We will address the Network Stability Analysis in a specific section.

Model and variable selection

Data's measurement level

Selecting the appropriate network model depends primarily on the data's measurement level. For cross-sectional data, the most common framework in psychology is the Pairwise Markov Random Field (PMRF), which models relationships between variables as conditional associations (Costantini et al., 2015; van Borkulo et al., 2014). Variables that are connected, conditional on all other variables in the data, are considered to be probabilistically dependent (Epskamp, Borsboom, & Fried, 2018). The specific measure of conditional association depends on the data measurement level. For ratio or interval level data, an appropriate model is the Gaussian Graphical Model (GGM), in which edges are parameterized as partial correlation coefficients (i.e. a partial correlation network) (Costantini et al., 2015; Lauritzen, 1996). GGMs are characterized as undirected, weighted graphs, meaning relationships are symmetrical (A ↔ B), and edges encode the magnitude of the association. The GGM also allows the use of polychoric correlations when the data are ordinal (Epskamp, 2016). For binary data, the Ising model is suggested, in which edges represent log-linear relationships (van Borkulo et al., 2014). Each model can be computed using a variety of estimators in R (R Core Team, 2021); see Figure 1 for an overview of available models and estimators.

Variable selection

Node selection in network psychometrics must be strongly theory-driven. Researchers should carefully determine which variables to include based on their specific research questions. A well-constructed network should incorporate all variables theoretically expected to be relevant to, or to influence, the clinical condition or cognitive system under investigation. Without a clear theoretical rationale for node selection, it remains ambiguous what scientific question the model is answering or how it advances our understanding of the underlying construct. For instance, the inclusion of non-cognitive measures, such as depression or sleep questionnaires, should be justified by explicit hypotheses regarding their interplay with cognitive domains. Employing network analysis without a defined aim risks yielding purely descriptive models that fail to address meaningful questions. As Fried (2020) emphasized, for the field to progress, it must transition from discovery-oriented research, merely establishing that nodes are connected, to theory-testing research, where explicit theoretical frameworks justify the inclusion of every variable in the model.

Beyond theoretical considerations, researchers must recognize the statistical implications of node selection. In a partial correlation network, each included variable adjusts the overall structure by partialling out the variance it captures. This is particularly crucial when incorporating global cognitive screeners (e.g. MoCA or MMSE), which naturally share widespread variance with multiple specific cognitive tests and can thereby structurally alter the network. Similarly, including highly overlapping tests from the same cognitive domain can deplete shared variance due to multicollinearity, potentially resulting in artificially weakened edges. Therefore, we strongly recommend checking for excessive statistical overlap prior to network estimation and removing redundant measures. Following this principle, when a single task yields multiple scores, researchers should avoid entering them simultaneously and instead opt for either a composite global score or specific, theoretically distinct subtest scores. Furthermore, researchers should prioritize variables with high measurement reliability, as the inclusion of noisy or highly unreliable variables can compromise the stability of the estimated network parameters.

The inclusion of demographic variables (e.g. age, education) also requires careful theoretical and methodological deliberation. Since including demographics directly in the network isolates the relationships among cognitive tests net of demographic influences, researchers should input raw, rather than demographically adjusted, cognitive scores to prevent overcorrection. Conversely, if the sample features a restricted age range or consists of multiple groups already matched on demographic characteristics, these variables can safely be omitted. Ultimately, the interpretation of any network model is fundamentally contingent upon the initial research questions, the specific variables selected, the shared variance among them and the theoretical framework guiding these choices.

The *neuropsychological_example* database includes performance on 11 neuropsychological tests (MMSE, FAB, RVLTi, RVLTd, DSf, DSb, PhonF, SemF, BNT, CDT and VSil). We are interested in assessing the organization and relationships between these tests. We do not include demographic information in the network because a previous paper (Tosi et al., 2020) found a marginal impact on neuropsychological performance in cases of cognitive impairment. Since our nodes are measured at an interval level (test scores) and we analyse a cross-sectional sample, we select a GGM estimated with the *estimateNetwork()* function of the R package *bootnet* (R lines 107–115), which allows for estimating networks and their stability in a generalized framework (Epskamp, Borsboom, & Fried, 2018).

Assumptions

A key benefit of network analysis is its lack of strong assumptions about the data-generating model, making it particularly suited to exploratory analysis (Borsboom et al., 2021). However, the validity of the results relies on several core assumptions (Burger et al., 2023).

The first assumption of the cross-sectional PMRF is the independence of observations (e.g. one distinct subject per row). Methodological Frontiers in Network Psychometrics section will cover more advanced methods applicable to longitudinal designs, where observations are not independent.

A second assumption is the presence of variables normally distributed, (log) linear relationships and pairwise interactions only. Methods to check for (multivariate) normality exist, and for continuous data that are not normally distributed, a transformation can be applied (e.g. non-paranormal transformation; Liu et al., 2012) before estimating the GGM (Epskamp, Borsboom, & Fried, 2018). This transformation relaxes the normality assumption by applying a smooth monotonic mapping to the data, ensuring the variables are normally distributed before the network is estimated. Normality can be violated in many ways, and the impact has not yet been investigated in detail in the context of GGMs (Epskamp, Borsboom, & Fried, 2018). As such, it is not yet clear how to proceed when normality is violated, making it crucial to use data-driven bootstrap methods to complement any analyses (Epskamp, 2020), as we will see in the Network Stability Analysis section.

A third assumption is that missing data are missing at random (MAR), meaning the probability of a value being missing is related to other observed data, but not to the missing value itself; or completely at random (MCAR), meaning the probability of missing data is unrelated to any observed or unobserved data (Rubin, 1976). To handle missing data, the *bootnet* package (Epskamp, Borsboom, & Fried, 2018), which we will use in our step-by-step tutorial, uses the average pairwise sample size as a proxy for the sample size in the GGM estimator; instead, the *psychonetrics* package (see Figure 1) includes full-information maximum likelihood estimation (Epskamp et al., 2022), which uses all observed data to estimate the network structure.

In our case study, we assume independent observations, since each row of the database corresponds to a single subject and assess the normality assumption (R lines 59–68). As shown in the histograms (Figure 2, upper panel), several variables, such as phonemic fluency test (PhonF), Rey Auditory Verbal Learning Test delayed recall (RVLTd) and incomplete letters subtest of the visual object and space perception battery (VSil), are skewed. Moreover, the multivariate normality test (R line 68) indicates significant deviations from normality (p-values $<.001$).

Given the violation of the normality assumption, we employ the non-paranormal transformation (R lines 75–77, Figure 2, lower panel). This ensures that the partial correlations are based on latent

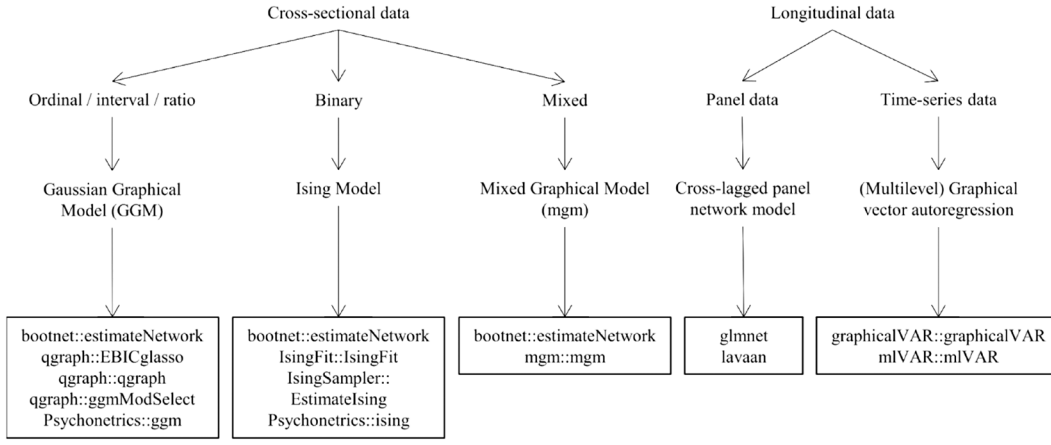

FIGURE 1 Decision tree. The figure provides an overview of the R packages and estimators (package::Estimator) most commonly used in neuropsychology, along with a decision tree to guide estimator selection based on the data design and the variables' measurement scales.

variables that satisfy the Gaussian assumption (R lines 79–85), providing more reliable edge estimates than using raw skewed data (Liu et al., 2012; Epskamp, Borsboom, & Fried, 2018). As the RVLTd still exhibits skewness (Figure 2, lower panel), we complement our analysis with data-driven bootstrap methods to assess the network stability (Epskamp, 2020). The database does not contain missing values (see Tosi et al., 2024, for a description of the missing data handling strategy).

Network structure estimation

Before executing the network estimation, it is best practice to initialize R's pseudorandom number generator using the *set.seed()* function (R line 105) in order to make results fully reproducible.

The GGM can be obtained using the *estimateNetwork()* function from the *bootnet* package and the *default* = *'EBICglasso'* argument (R lines 107–115). This method computes a sparse inverse covariance matrix and uses the Extended Bayesian Information Criterion (EBIC) to select the optimal regularization parameter (see the Edge Selection and Pruning section for details). The estimated GGM is represented as a square matrix (p × p) where each row and column corresponds to a node (i.e. a variable) in the network. The elements of the matrix represent partial correlation coefficients, indicating the strength and sign of the connection between two nodes (R lines 129–132).

Edge selection and pruning

In GGMs, two nodes are connected if their connection (i.e. partial correlation) is non-zero. When calculated from empirical data, these partial correlations will almost never be exactly zero due to sampling variation (Costantini et al., 2015; Hevey, 2018). Consequently, without intervention, a raw network estimate will be fully connected (i.e. every node is linked to every other node). Although in neuropsychology, performance on cognitive tasks, as measured by neuropsychological tests, exhibits a positive manifold (i.e. universal positive covariance; Van Der Maas et al., 2006), a fully connected network is hardly a true representation of the underlying cognitive architecture, and failing to account for this can lead to wrong conclusions. Moreover, in neuropsychology, we often face the challenge of small sample sizes (Epskamp, Borsboom, & Fried, 2018; C. Ferguson, 2021; Tosi et al., 2020; Tosi, Nigro, et al., 2024). Indeed, the more nodes in the network, the more edges need to be estimated, and the more observations

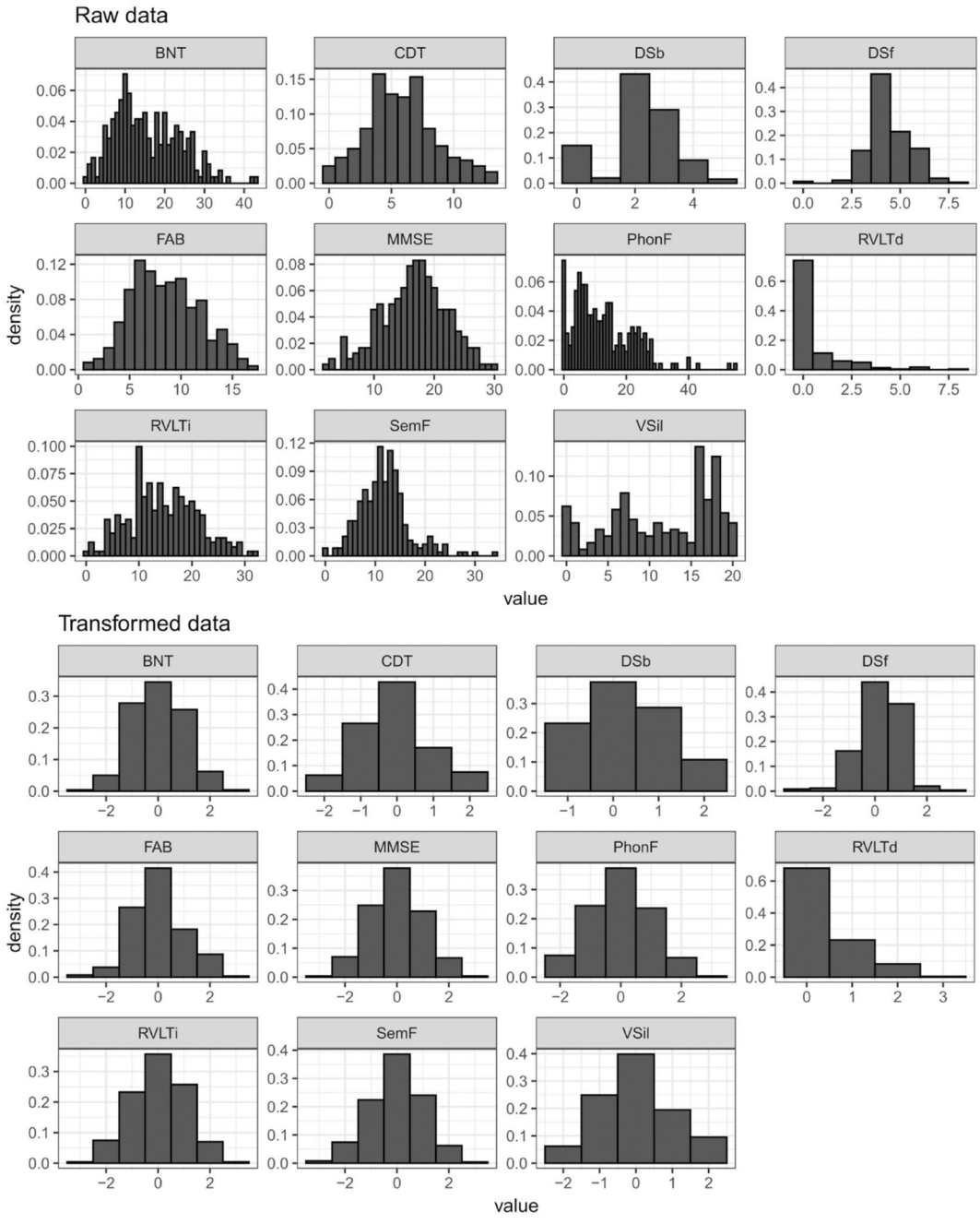

FIGURE 2 Histograms. The upper panel shows the raw data distributions, the lower panel shows the transformed data distributions. BNT, Boston Naming Test; CDT, Clock Drawing Test; DSb, digit span backward; DSf, digit span forward; FAB, Frontal Assessment battery; MMSE, Mini Mental State Examination; PhonF, phonological fluency test; RAVLi, Rey Auditory Verbal Learning Test—immediate recall; RAVLd, Rey Auditory Verbal Learning Test—delayed recall; SemF, Semantic Fluency test; VSil, incomplete letters subtest of the visual object and space perception battery.

are needed. In such scenarios, the model captures not only the true signal (genuine cognitive interactions) but also the random noise specific to that sample, which may lead to overfitting and unstable estimates (Babyak, 2004).

In general, dense networks and spurious relationships can be problematic for the interpretation and replicability of the results. On the contrary, an optimal GGM is sparse (i.e. with many absent edges) and maintains a high likelihood (Costantini et al., 2015). To this end, edge selection methods are frequently used. The literature presents multiple methods (i.e. null-hypothesis testing, cross-validation and regularized estimation); however, one emerged as predominant for cross-sectional GGM. Regularized estimation procedures aim to include only edges that improve the fit of the network model to the data, thereby balancing parsimony and fit (Borsboom et al., 2021). This is the preferred method for estimating psychological networks in moderate sample sizes because it effectively prunes the network while balancing goodness-of-fit with parsimony (Costantini et al., 2015).

The most used regularization technique is the least absolute shrinkage and selection operator (LASSO; Tibshirani, 2011). The LASSO technique applies a penalty to the model that limits the total sum of absolute parameter values and causes estimates of weak connections to shrink toward zero. Consequently, the LASSO yields a sparse (or parsimonious) network, in which only a relatively small number of edges are retained to explain the covariation structure of the data. Sparsity enhances the network's interpretability and significantly reduces the false positives rate (Costantini et al., 2015; Hevey, 2018). The degree of sparsity is controlled by a tuning parameter (λ). The optimal value of λ is chosen based on the Extended Bayesian Information Criterion (EBIC; Chen & Chen, 2008). This method has been shown to perform well in recovering the true network structure for both Ising and GGM models (Foygel & Drton, 2010; van Borkulo et al., 2014). In turn, the EBIC calculation relies on a hyperparameter, γ, which defines how strongly the criterion prefers simpler models. The γ ranges from 0 to 1. A value of 0 results in a dense network, potentially including more false positives. The default setting in most packages is a conservative approach (i.e. $\gamma = .5$), which prioritizes specificity and ensures that the edges retained are likely to be true. On the other hand, this choice may result in the omission of some true, albeit weak, relationships (Hevey, 2018). Although every value of gamma is possible, standard values should be adopted unless strong motivations emerge: .25 (less conservative), .5 (conservative, default) and 1 (very conservative) (Ávalos-Tejeda & Calderón, 2025). These default settings have consistently proven satisfactory in balancing sensitivity and specificity in applied psychological research (Epskamp, 2016).

In our tutorial script, we use the default γ value of .5 (R line 110) and test 100 λ tuning parameters (R line 112). We set the minimum λ ratio (i.e. the lower bound of the λ value to be tested) to .01, thus fixing the minimum penalty to 1% of its maximum value (R line 111). Since the LASSO estimation is an iterative procedure, it can sometimes produce non-zero edge weights that are infinitesimally small (e.g. 10^{-7}). We force these negligible values to exact zeros, ensuring a truly sparse network structure and improving the model's specificity (R line 115).

Visualization and results

Once the network is estimated, we can access the inverse covariance matrix, which encodes partial correlation coefficients between variables. The matrix is stored in the *graph* value of the saved network (R lines 130). Table 1 shows the inverse covariance matrix, where each row and column corresponds to a network node, and each element of the matrix represents the strength and sign of the connection between the two nodes.

The most powerful way to inspect the resulting structure is to visualize the network itself with a plot. We can use the *plot()* function (R lines 145–146) or customize the network using the *qgraph()* function from the *qgraph* package (R lines 148–154).

The default settings of both functions use the Fruchterman-Reingold algorithm to visualize the network (Fruchterman & Reingold, 1991). This force-directed algorithm positions nodes based on the strength of their connections: nodes with stronger associations attract each other, while all nodes exercise a repulsive force to prevent overlap. Consequently, highly connected nodes are visualized closer and centrally, while isolated nodes are pushed to the periphery (Figure 3, panel a). Edges in the network (Figure 3, panel a) represent partial correlations as expressed in the inverse covariance matrix (Table 1). The absence

TABLE 1 Inverse covariance matrix.

	MMSE	FAB	RVLTi	RVLTd	DSf	DSb	PhonF	SemF	BNT	CDT	VSil
MMSE	.000	.173	.000	.000	.000	.160	.140	.145	.155	.195	.000
FAB	.173	.000	.000	.000	.000	.176	.338	.000	.000	.000	.266
RVLTi	.000	.000	.000	.388	.158	.000	.000	.159	.270	.000	.000
RVLTd	.000	.000	.388	.000	.000	.000	.000	.000	.000	.000	.000
DSf	.000	.000	.158	.000	.000	.174	.209	.000	.000	.000	.000
DSb	.160	.176	.000	.000	.174	.000	.000	.000	.000	.000	.000
PhonF	.140	.338	.000	.000	.209	.000	.000	.000	.169	.000	.000
SemF	.145	.000	.159	.000	.000	.000	.000	.000	.000	.000	.000
BNT	.155	.000	.270	.000	.000	.000	.169	.000	.000	.000	.000
CDT	.195	.000	.000	.000	.000	.000	.000	.000	.000	.000	.227
VSil	.000	.266	.000	.000	.000	.000	.000	.000	.000	.227	.000

Note: Each row and column corresponds to a network node, and each matrix element represents the strength and sign of the connection between the two nodes.

Abbreviations: BNT, Boston Naming Test; CDT, Clock Drawing Test; DSb, digit span backward; DSf, digit span forward; FAB, Frontal Assessment battery; MMSE, Mini Mental State Examination; PhonF, phonological fluency test; RAVLi, Rey Auditory Verbal Learning Test—immediate recall; RAVLd, Rey Auditory Verbal Learning Test—delayed recall; SemF, Semantic Fluency test; VSil, incomplete letters subtest of the visual object and space perception battery.

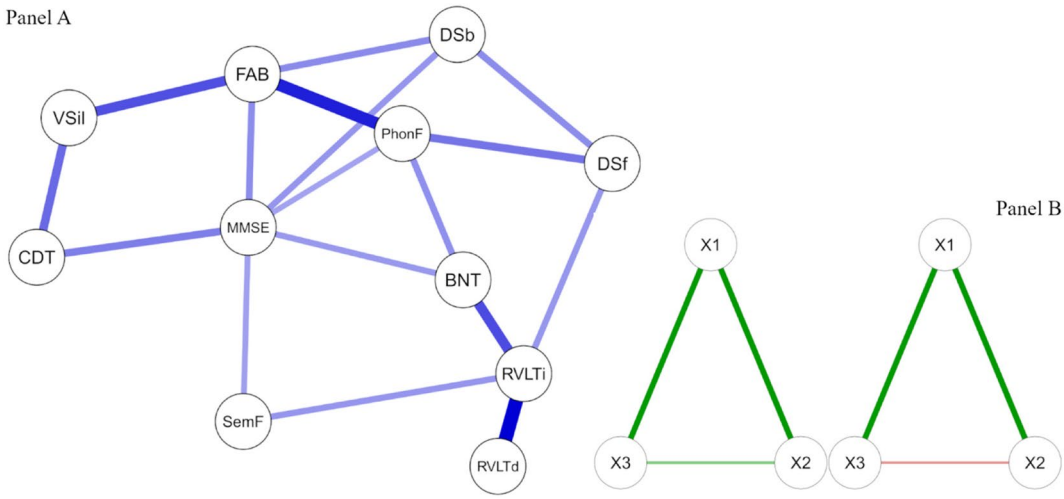

FIGURE 3 Network plot. Panel a shows the network analysis result. Nodes represent neuropsychological variables; edges represent their relationships. Blue edges show positive relations. The edges' size and colour saturation represent the relationships' intensity. BNT, Boston Naming Test; CDT, Clock Drawing Test; DSb, digit span backward; DSf, digit span forward; FAB, Frontal Assessment battery; MMSE, Mini Mental State Examination; PhonF, phonological fluency test; RAVLi, Rey Auditory Verbal Learning Test—immediate recall; RAVLd, Rey Auditory Verbal Learning Test—delayed recall; SemF, Semantic Fluency test; VSil, incomplete letters subtest of the visual object and space perception battery. Panel b shows a collider example: The left graph shows two variables (x2 and x3) that both positively contribute to a third variable (x1; the collider). Controlling for that third variable within the network (i.e. the right graph) statistically forces a negative partial correlation between the two variables.

of an edge between two nodes implies conditional independence, suggesting no unique relationship exists between them after controlling for all other variables in the network. For example, Table 1 shows conditional independence between the digit span forward (DSf) and the Clock Drawing Test (CDT); consequently, Figure 3, panel a, shows no edges between these nodes. Crucially, the absence of an edge

does not prove the relationship is zero, but rather that the current data do not provide sufficient evidence to include it in the regularized model (Costantini et al., 2015). On the other hand, the presence of an edge indicates a statistical dependency between the two nodes, but it does not specify the nature of that relationship. An edge could represent a direct causal effect, a reciprocal interaction, an overlap in the variable content, or the influence of an unmeasured latent variable. For example, Figure 3, panel a, shows an edge between SemF and RAVLTi, suggesting that, controlling for all other variables, there is a unique relationship between the semantic fluency task and the immediate memory recall subtest.

The edge colour indicates the sign of the relationship: blue (or green) edges represent positive partial correlations (where an increase in one node is associated with an increase in the other), while red edges represent negative partial correlations (inhibitory or competitive relationships). In our illustrative network, we observe only positive partial correlations, visualized as blue edges. This indicates positive relationships, suggesting that people who score well on one cognitive test are likely to score well on the others (i.e. the positive manifold; Van Der Maas et al., 2006). Although we do not observe any negative edges in our current example, researchers often encounter a triplet in which two nodes are positively connected to a third node but negatively connected to each other (Figure 3, panel b). This counterintuitive negative edge often arises from conditioning on a collider. If two cognitive skills both positively contribute to a third variable (the collider), controlling for that third variable within the network statistically forces a negative partial correlation between the two skills. Researchers can often spot this artefact by checking the simple (zero-order) correlation between the two variables (R lines 134–135), which will typically be positive or zero despite the negative edge in the graph.

The colour saturation and width of an edge are proportional to the strength of the partial correlation: more saturated, thicker edges indicate stronger associations. When the network is visualized with *qgraph()* (Epskamp et al., 2012), the colour saturation and width are adjusted to the absolute weight and scaled relative to the graph's strongest weight. For example, we can see in Figure 3, panel a, that the edge between the Boston Naming Test (BNT) and the phonological fluency test (PhonF) is smaller than the edge between the phonological fluency test (PhonF) and the Frontal Assessment Battery (FAB). This information aligns with the strength of the connections shown in Table 1, where the relationship between BNT and PhonF (.169) is lower than that between PhonF and FAB (.338). We can interpret this information as indicating that phonological fluency is more closely related to frontal executive functions, as assessed by the FAB, than to naming skills, as assessed by the BNT.

Network description and metrics

Once a network is estimated, several indices can be computed. The *print()* function (R line 164–165) shows some descriptors of the estimated network, such as the number of nodes (i.e. variables), the number of non-zero edges on the total of possible edges in the network and the mean weight (i.e. the average value of the strengths assigned to the edges in the network).

The number of estimated (non-zero) edges relative to the total number of possible edges (17/55) is the network density, which is .309 in our example and indicates how dense versus sparse the network is. A dense network refers to a network structure with many connections, whereas a sparse network refers to a structure with few connections. The network mean weight (.064) is calculated as the average of the absolute weights of all edges in the network, providing a summary statistic of overall system connectivity. While this metric can offer preliminary insights into the degree of global cognitive integration (i.e. higher values might suggest interdependence and lower values a more modular architecture), we caution against overinterpreting it in isolation. The network mean weight must be interpreted alongside other metrics (i.e. network density) and the specific structural distribution of edges. Furthermore, a high mean weight does not necessarily imply genuine cognitive integration; it can sometimes reflect mutual method variance, such as shared stimuli, instructions, or scoring methods among the administered tests. Consequently, we outline its utility primarily as a comparative index between networks (e.g. using the Network Comparison Test; van Borkulo et al., 2023b; see Methodological Frontiers in Network Psychometrics section). In these

comparative scenarios, evaluating differences in mean weight is highly informative, provided that the set of variables and methodological influences remains constant across the groups being compared.

Focusing on the network's nodes, not all are equally important (Costantini et al., 2015; Hevey, 2018). In graph theory, their relative influence is quantified using centrality indices. The most common indices include:

Node Strength (and Degree)

Degree is the number of edges connected to a node. In weighted networks like the GGM, this is generalized to node strength, defined as the sum of the absolute weights of all edges connected to a specific node. A node with high strength is, on average, strongly conditionally associated with many other variables in the network. This is generally considered the most robust and interpretable centrality metric in psychological networks. Node strength divided by degree allows for the evaluation of a node's mean weight, which is the average intensity of the node's connections. The node's mean weight allows researchers to differentiate between nodes that exert diffuse influence through many weak connections and those that rely on a few highly robust connections.

Expected Influence

A limitation of Node Strength is that it uses absolute values, treating positive and negative edges equally. Expected Influence (Robinaugh et al., 2016) sums the raw edge weights. This is particularly useful in networks containing negative edges, as it indicates the directional cumulative impact a node might have on the network.

Closeness and Betweenness

Closeness quantifies the inverse sum of the shortest distances from a node to all other nodes, while betweenness quantifies how often a node lies on the shortest path between any two other nodes. Recent methodological guidelines strongly advise against using closeness and betweenness in psychological cross-sectional networks (Borsboom et al., 2021; Bringmann et al., 2019). These metrics treat associations as physical distances. In a PMRF, where edges represent partial correlations and paths convey no information transfer, this assumption is violated, rendering closeness and betweenness highly unstable and conceptually difficult to interpret.

Predictability

It quantifies the absolute variance of a node that is explained by its direct neighbours (analogous to R^2 in multiple regression). This indicates how well the network structure accounts for the state of that specific variable.

Bridge Centrality

If a network contains distinct clusters or communities (e.g. a 'Memory' cluster and an 'Executive Function' cluster), bridge centrality identifies nodes that strongly connect these different domains, acting as pathways between symptom clusters (Jones et al., 2021).

The *qgraph* package allows for the estimation and visualization of centrality indices via the *centralityTable()* and *centralityPlot()* functions (R lines 180–201). In our illustrative example, we will focus on strength, since we have no negative edges, and it equals the expected influence. As highlighted before, closeness and betweenness are unstable and conceptually difficult to interpret. Figure 4 shows the standardized strength index for each edge.

Delayed memory recall, the MMSE and the FAB are the strongest nodes in the network (Figure 4). This result suggests that screening tests play an integrative role, which is informative for global systemic assessment but distinct from the specific diagnostic value of domain-specific tests. Indeed, their connectivity patterns differ significantly when accounting for mean weight: the MMSE is connected to the network via many weak edges, typical of a non-specific global screening test (mean weight = .161). In contrast, the executive (FAB; mean weight = .238) and memory (RAVLTd; mean weight = .388) tests exhibit fewer, stronger edges. This structural difference is consistent with clinical theory, in which executive functions exert robust top-down regulation over specific cognitive domains (Ferguson & Foley, 2024) and delayed recall serves as the core, specialized deficit in Alzheimer's Disease.

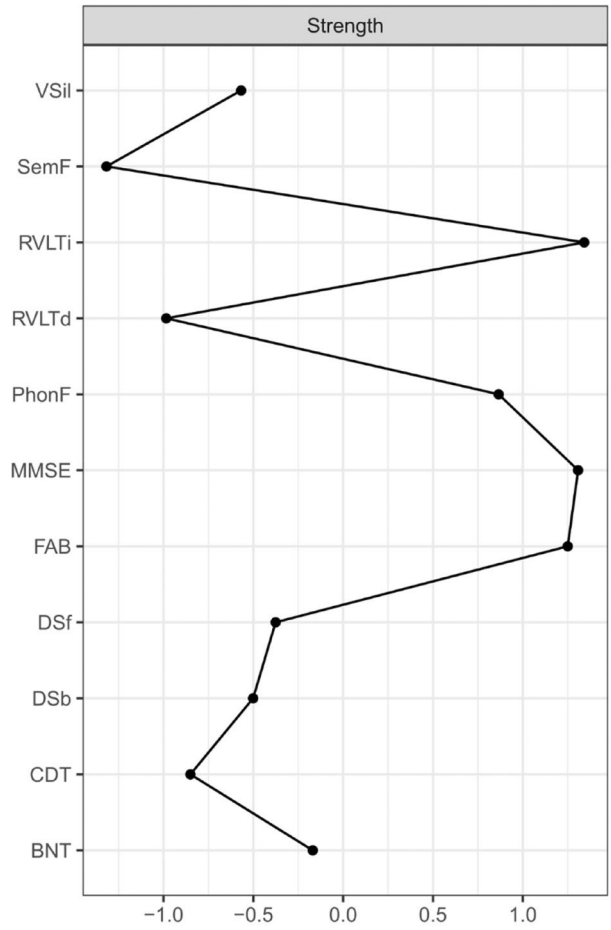

FIGURE 4 Centrality indices. The figure shows the network nodes on the *y*-axis and standardized strength scores on the *x*-axis. BNT, Boston Naming Test; CDT, Clock Drawing Test; DSb, digit span backward; DSf, digit span forward; FAB, Frontal Assessment battery; MMSE, Mini Mental State Examination; PhonF, phonological fluency test; RAVLi, Rey Auditory Verbal Learning Test—immediate recall; RAVLd, Rey Auditory Verbal Learning Test—delayed recall; SemF, Semantic Fluency test; VSil, incomplete letters subtest of the visual object and space perception battery.

Network stability analysis

Because psychological networks are estimated from empirical sample data, edge weights and centrality indices are subject to sampling variability (Hevey, 2018). Particularly in studies with moderate sample sizes, these estimates may be unstable. Therefore, estimating a network is only the first step; researchers should always evaluate the accuracy of the edge weights and the stability of the node centralities (Epskamp, Borsboom, & Fried, 2018). The *bootnet* package provides a comprehensive framework to perform stability analysis via resampling techniques (Epskamp, Borsboom, & Fried, 2018).

Edge-weight accuracy

To evaluate the precision of the estimated edge weights, we can construct a sampling distribution of each edge and compute a 95% Confidence Interval (CI) around it. The regularized partial correlations distribution is derived through bootstrapping (Efron, 1979). For LASSO-regularized networks, the non-parametric bootstrap is highly recommended (Epskamp, Borsboom, & Fried, 2018; Hevey, 2018). In this data-driven approach, observations are repeatedly resampled with replacement to create thousands of plausible datasets, and the network is re-estimated for each. This generates a distribution of weights for every edge, from which a 95% CI is derived. Narrow CIs indicate that the edge weight was estimated with high precision. Wide CIs indicate poor precision, meaning the exact strength of the connection is uncertain. Edge CIs should be used only to assess the precision of the estimate and to compare the relative strength of different edges, not to test their existence (Epskamp, Borsboom, & Fried, 2018). Indeed, the LASSO penalty has already performed model selection: if an edge is present in the estimated network, it is sufficiently relevant to be retained. Moreover, the overlap of bootstrap estimates and sample observations can be used to assess stability (see the example below for a clear interpretation).

Lines 218–230 of the R script generate bootstrapped replicas and CIs to assess the edge-weight accuracy in our case study, using the *bootnet()* function. Results can be inspected via a table (R lines 232–234) or a plot (R lines 236–240), which we show in Figure 5. Each horizontal line represents one edge of the network, ordered from the highest to the lowest edge weight. The red line indicates the sample values, and the black line represents the bootstrap mean value; the more similar these values are, the more robust the results are. We observe only slight differences in a limited number of edges. The minor divergence between the sample estimates and bootstrap means for the weaker edges (i.e. near 0) reflects the expected shrinkage effect of the LASSO regularization across bootstrap samples. The grey area represents the bootstrapped CIs; the narrower the CIs, the more replicable the results. Overall, the confidence intervals are moderately wide, suggesting some uncertainty about the exact magnitudes of the partial correlations. However, the order of the strongest edges is highly reliable. For instance, the edges connecting FAB to PhonF, and RVLTi to RVLTd, are distinctly stronger than the majority of the network, as their lower confidence bounds do not substantially overlap with the point estimates of the remaining edges. Following current methodological guidelines, we do not use these intervals to test for significant differences from zero, but rather to confirm that the largest edge weights in our network are robustly estimated.

Centrality stability

To assess the reliability of node centrality, Epskamp et al. (2018) developed the case-dropping subset bootstrap. This technique assesses whether the rank order of the centrality indices remains stable when portions of the data are systematically removed. The network is re-estimated using decreasing subsets of the original sample (e.g. retaining 90%, 80%, 70%, down to 25% of cases). For each subset, the correlation between the new centrality indices and the original full-sample

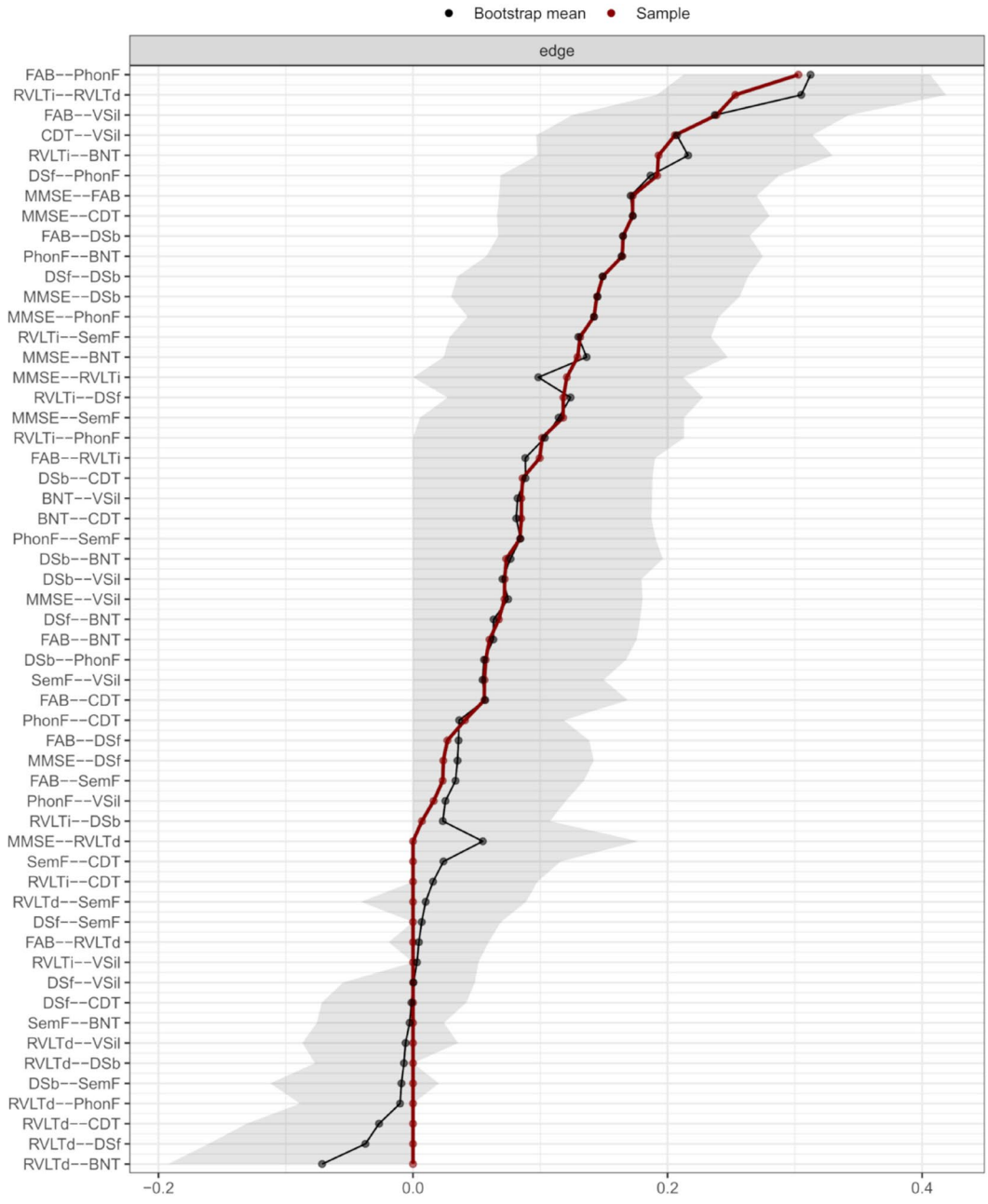

FIGURE 5 Edge-weight accuracy. Each horizontal line represents one edge of the network, ordered from the highest to the lowest edge weight. The red line represents the sample value, the black line represents the bootstrapped mean value. Grey areas represent bootstrap CIs. BNT, Boston Naming Test; CDT, Clock Drawing Test; DSb, digit span backward; DSf, digit span forward; FAB, Frontal Assessment battery; MMSE, Mini Mental State Examination; PhonF, phonological fluency test; RAVLi, Rey Auditory Verbal Learning Test—immediate recall; RAVLd, Rey Auditory Verbal Learning Test—delayed recall; SemF, Semantic Fluency test; VSil, incomplete letters subtest of the visual object and space perception battery.

indices is calculated. To quantify the stability, researchers can use the correlation-stability (CS) Coefficient. The CS coefficient (typically calculated with a target correlation of .7) represents the maximum proportion of cases that can be dropped from the dataset such that, with 95% probability, the correlation between the subset centralities and the original centralities remains at least .7.

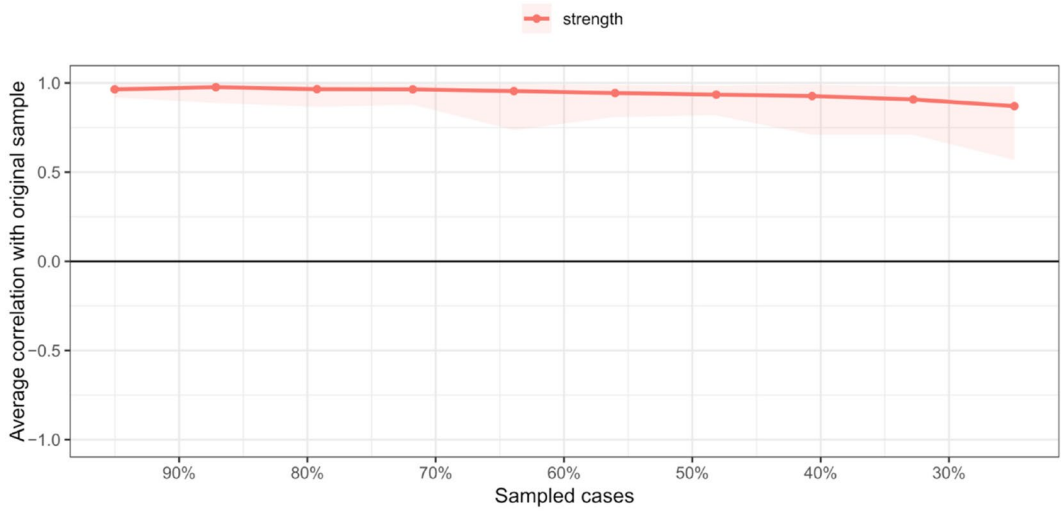

FIGURE 6 Centrality stability. The plot shows the average correlations between the strength of networks sampled with persons dropped and the original sample on the *y*-axis and the percentage of sample cases on the *x*-axis. The line indicates the mean correlation, and the coloured area shows the range from the 2.5th to the 97.5th quantiles.

The CS coefficient should not fall below .25, and it is highly preferable for it to be >.50 (Epskamp, Borsboom, & Fried, 2018).

For centrality stability, we use the *bootnet()* function and compute the case-dropping bootstrap (R lines 253–263). The bootstrapped centrality plot (Figure 6; R lines 268–269) shows the average correlations between centrality indices computed in networks sampled with persons dropped and the original sample. For multiple centrality indices, each index is shown as a line of a different colour; lines indicate the means, and areas indicate the range from the 2.5th quantile to the 97.5th quantile. We can see that the correlations remain high even when we retain only 30% of the original sample. The same results can be inspected in a table (R lines 265–266).

The correlation-stability (CS) coefficient can be computed with the *corStability()* function (R lines 271–274). The analysis reveals a CS coefficient of .672, meaning that we can randomly drop up to 67.2% of our sample, and 95% of the time, the rank order of the node strengths in that smaller subset will still correlate highly ($r = .70$) with the original network. Because this value exceeds the recommended strict threshold of .50 (and is well above the minimum acceptable threshold of .25; Epskamp, Borsboom, & Fried, 2018), we can conclude that the rank order of node centrality in our network is highly stable.

Significant Differences

Beyond assessing absolute accuracy, the bootstrapped distributions can be used to test for significant differences between specific network parameters. Using the bootstrapped difference test (*differenceTest()* function) included in the *bootnet* package, researchers can evaluate whether the weight of edge A is significantly different from the weight of edge B, or if the centrality of node X is significantly higher than that of node Y.

Wrap up: What could we discuss in our example scenario?

To sum up, we estimated a GGM of 11 neuropsychological tests. The analysis revealed unique relationships between cognitive tests, such as the diffuse connections of the MMSE, which proved to represent a variety of cognitive performance and to play an integrative role. The most relevant (central) nodes

were the RVLTi, MMSE and FAB, suggesting that screening and memory tests are the most informative in this neuropsychological evaluation, although with distinct roles. Finally, we can consider our results reliable and stable, given the stability coefficients and the edge CI.

Network analysis in Jasp: A user-friendly alternative

R is not the only software that enables network analyses. For researchers who prefer a point-and-click interface over command-line programming, JASP (an open-source statistical software suite; JASP Team, 2025) offers a dedicated, highly intuitive 'Network' module. Because JASP's network module is built directly on top of the R packages discussed in this tutorial (such as *bootnet* and *qgraph*), users can conduct rigorous analyses without writing any code.

Key capabilities available in JASP include:

Model Selection

You can easily estimate GGMs using the *EBICglasso* estimator for continuous or ordinal data. The module also supports Ising models (*IsingFit*) for binary data and Mixed Graphical Models (*mgm*) for datasets with varying measurement levels.

Network Description

The software generates customizable network plots, computes and visualizes core centrality indices, including Node Strength, Closeness, Betweenness, and Expected Influence.

Accuracy and Stability

The software integrates the bootstrapping procedures outlined in this tutorial. Researchers can perform non-parametric bootstrapping to obtain confidence intervals for edge-weight accuracy and case-dropping bootstrapping for centrality indices.

While JASP is an excellent tool for standard analyses and rapid exploration, R remains the most powerful environment due to its flexibility. JASP relies primarily on default settings. In contrast, writing scripts in R allows researchers to fine-tune every parameter of the estimation process. For instance, if an anomaly occurs during estimation (such as the lowest lambda warning discussed earlier) in R, we can adjust the minimum lambda ratio. Furthermore, R provides total control over data visualization and reporting. Using the full capabilities of *qgraph* and *ggplot2*, researchers can deeply personalize network plots and generate highly customized, publication-ready tables that journal formats often require. Ultimately, while JASP lowers the barrier to entry, R provides the absolute control necessary for complex, highly customized, or troubleshooting-intensive network psychometrics.

METHODOLOGICAL FRONTIERS IN NETWORK PSYCHOMETRICS

The cross-sectional GGM described in How to Create a Neuropsychological Network: A Step-by-Step Guide section represents a useful starting point for estimating conditional associations among neuropsychological test scores. However, many neuropsychological research questions require extensions of this basic model. These approaches are not inherently superior to basic models; rather, each addresses

a specific research question or limitation inherent in cross-sectional estimation. Understanding these frontiers allows neuropsychologists to align methodological choices with substantive aims.

Identifying cognitive dimensionality: Exploratory graph analysis

One important development concerns dimensionality. Neuropsychology has traditionally relied on factor-analytic techniques to identify latent cognitive domains. However, network theory has challenged the strict opposition between latent and network perspectives. Under certain conditions, clusters of strongly connected nodes in a network closely correspond to latent factors (Briganti et al., 2024). Exploratory Graph Analysis (EGA) leverages this insight by applying community detection algorithms to estimated networks in order to identify empirically emergent dimensions (Golino & Epskamp, 2017). Rather than extracting factors through eigenvalue decomposition, EGA detects strongly interconnected nodes within the network. For neuropsychological batteries, this approach may provide a complementary method for evaluating whether assumed domains—such as executive functioning or visuospatial ability—emerge naturally from observed interdependencies (Tosi, Nigro, et al., 2024). Importantly, EGA does not eliminate the possibility of latent constructs; instead, it offers a structurally grounded route to dimensional discovery.

Notably, EGA has been extended to longitudinal panel data, allowing investigation of cognitive dimensionality over repeated measures designs (Golino et al., 2022).

Comparing cognitive organization across clinical groups: Joint graphical models and network comparison tests

Another frontier concerns formal group comparisons. Neuropsychological research frequently contrasts clinical populations—such as Alzheimer's disease, vascular dementia, Parkinson's disease, or mild cognitive impairment—with healthy controls. Traditional analyses focus on mean differences or regression-based predictors. Network approaches allow structural comparisons through procedures such as the Network Comparison Test (NCT), which evaluates differences in global strength, specific edge weights, or overall structure between groups (van Borkulo et al., 2023). In contexts where pathology is hypothesized to involve cognitive reorganization rather than mere performance decline, such tests offer a way to quantify architectural change (Longo et al., 2024; Tosi, Nigro, et al., 2024). However, these procedures require adequate sample sizes and careful control of estimation parameters, as differences may otherwise reflect sampling variability rather than genuine structural divergence.

A complementary strategy involves estimating networks jointly across groups rather than separately. Methods based on the joint graphical LASSO implement this idea by simultaneously estimating multiple networks while encouraging shared structure between them (Danaher et al., 2014). In practice, this approach introduces regularization terms that both promote sparsity within each network and penalize large discrepancies across groups. Implementations such as the *estimateGroupNetwork()* function allow researchers to obtain group-specific networks that borrow statistical strength from one another while still permitting meaningful differences to emerge. This strategy can be particularly useful when sample sizes are modest—a common situation in neuropsychology—because the shared estimation framework stabilizes edge estimates and reduces the likelihood that group differences reflect estimation noise. Conceptually, joint estimation shifts the comparison problem from a purely post hoc procedure to a model-based one. Rather than first estimating networks independently and then testing their differences, the joint graphical LASSO treats group structure as part of the estimation process itself. As a result, similarities and differences between networks are inferred within a unified framework, which may provide more stable estimates of shared and group-specific associations. Nevertheless, interpretation still requires caution. Differences in estimated networks may arise from heterogeneity in variance structure, measurement reliability or distributional properties across groups and therefore should be considered alongside traditional neuropsychological evidence.

Modelling heterogeneous neuropsychological data: Mixed graphical models

Neuropsychological datasets often include heterogeneous variable types: continuous test scores, ordinal ratings, binary impairment indicators, and demographic covariates. Mixed Graphical Models (MGMs) extend the PMRF framework to accommodate such mixed data structures (Altenbuchinger et al., 2020). By parameterizing edges through generalized linear models, MGMs allow continuous, binary, and categorical variables to coexist within a single network. This flexibility is particularly relevant for translational settings, where diagnostic classifications and continuous performance metrics interact within the same assessment context. Moreover, *mgm* (Haslbeck & Waldorp, 2015) can accommodate a grouping variable as a categorical moderator and estimate group differences as moderation effects (Haslbeck & Waldorp 2020).

Studying change over time: Longitudinal and temporal network models

Perhaps the most conceptually transformative extensions involve longitudinal modelling. Cross-sectional networks capture between-person associations; they do not describe intra-individual dynamics. However, for questions concerning cognitive decline, compensation, or progression, temporal structure is essential (Figure 1). Depending on the sampling frequency, longitudinal network models generally fall into two distinct categories: intensive time-series models and panel data models. When data are collected from a single individual or a few individuals over many repeated time points (typically $T > 50$), we refer to intensive longitudinal data or time-series data. A typical example is Ecological Momentary Assessment (EMA), which measures experiences, such as personality states, symptoms, and behaviours, at the moment they occur, multiple times a day, thus capturing natural fluctuations over time. Unlike cross-sectional networks, these models allow us to estimate temporal (lagged) effects to unveil how well a variable at time t predicts a variable at time $t+1$ (Epskamp, 2020; Epskamp, Waldorp, et al., 2018). Temporal associations are typically modelled using vector autoregressive (VAR) models, which are regression models based on previous measurement occasions (Epskamp, Waldorp, et al., 2018). The most widely used estimators are the *mlVAR* and the *graphicalVAR* that produce separate network structures to characterize the dependence relations: directed (i.e. relationships between nodes are asymmetrical), temporal networks account for the association through time, undirected contemporaneous networks assess the associations between variables after temporal effects have been taken into account, and undirected between-person networks characterize the relationship between the stationary means of different subjects (Epskamp, Waldorp, et al., 2018). These algorithms assume that the time difference between measurements is roughly constant, the minimum number of measurements for each unit is 20, and local stationarity (i.e. the parameters do not change over time during the measurement period) holds (Blanchard et al., 2023; Epskamp, 2020; Epskamp, Waldorp, et al., 2018). While VAR assumptions are often met in daily EMA studies, they are highly unlikely to hold in neuropsychological evaluations or psychotherapy research that unfold over months or years (Falkenström et al., 2020; Romano et al., 2026). In these scenarios, researchers typically measure a large number of subjects at a few discrete time points (e.g. pre- and post-treatment, or annual follow-ups). This requires Panel Data approaches, which represent longitudinal relationships among variables and how their conditional associations change over time without requiring the strict stationarity assumption (Epskamp, 2020; Wysocki et al., 2022). As such, panel networks are particularly useful for tracking disease progression (e.g. Alzheimer's trajectories) or treatment mechanisms, allowing researchers to disentangle between-subject differences from within-subject changes over time. The Cross-Lagged Panel Network Model does not require the stationarity assumption and only expects that the sample size exceeds the total number of variables at each measurement occasion (Wysocki et al., 2022). In this model, cross-lagged effects are estimated by regressing a set of variables at a measurement occasion on those at the previous occasion, thereby obtaining the effect of each variable on the others at a particular time lag, while controlling for each variable's autoregressive effect on itself. Temporal associations are described as a directed network.

Ultimately, longitudinal and panel approaches disentangle stable individual differences from dynamic processes and offer a path toward modelling cognitive trajectories rather than static architectures. However, they require repeated measures designs and substantially larger data demands, and their interpretation becomes correspondingly more complex.

Quantifying uncertainty in specific theoretical relations: Bayesian network approaches

Finally, Bayesian approaches to network estimation are gaining prominence (Huth et al., 2023). Frequentist regularization techniques, such as the graphical LASSO, produce sparse point estimates but provide limited quantification of structural uncertainty. Bayesian methods allow estimation of posterior distributions over edges and, in some implementations, over entire network structures. This probabilistic framing is particularly attractive in neuropsychology, where moderate sample sizes and high-dimensional batteries are common. By explicitly modelling uncertainty, Bayesian approaches can temper overconfidence in single estimated networks and encourage more cautious inference.

Bayesian approaches are especially useful when the research question concerns the credibility of specific theoretically motivated edges rather than the overall visual structure of the network. For example, researchers may ask whether memory performance is conditionally associated with executive functioning after controlling for language and attention, whether a global screener retains unique associations with domain-specific tests, or whether a clinically relevant symptom such as depression, apathy, or fatigue is credibly connected to specific cognitive domains. In these cases, Bayesian estimation can quantify the uncertainty surrounding a given edge, helping researchers distinguish robust theoretical relations from associations that are too uncertain to interpret confidently. Moreover, Bayesian network analyses are also increasingly accessible to applied researchers, including through user-friendly software such as JASP.

Taken together, these developments illustrate the expanding methodological landscape of network psychometrics. The key challenge for neuropsychologists is not mastering every technique, but selecting the appropriate model for the research question at hand.

STRENGTHS, LIMITATIONS, AND RESPONSIBLE INTERPRETATION

Network psychometrics offers several advantages for neuropsychological research. First, it aligns naturally with the systemic nature of cognitive functioning. By modelling conditional dependencies among test scores, it reflects the interactive architecture that clinicians implicitly consider during assessment. Second, the conditional independence framework provides information beyond zero-order correlations, clarifying which associations remain after accounting for the broader system. Third, network visualization facilitates communication of complex multivariate structures, often revealing patterns that may remain obscured in tabular outputs. Finally, structural comparison methods allow formal testing of cognitive reorganization hypotheses, extending beyond mean-level differences.

These strengths, however, must be balanced against substantial limitations. The most fundamental concern is causal interpretation. Undirected Gaussian Graphical Models represent conditional associations, not causal pathways. An edge between two nodes indicates statistical dependence after controlling for other variables, but it does not establish directional influence. Cross-sectional networks are therefore descriptive representations of between-person structure. Causal claims require longitudinal designs, experimental manipulation, or stronger modelling assumptions.

A second critical issue concerns stability. Psychometric networks are high-dimensional statistical objects. In modest samples, edge weights and centrality metrics may fluctuate considerably across bootstrap resamples. Without systematic stability analysis, networks risk reflecting sampling noise rather

than structural signal. Bootstrap confidence intervals for edges and centrality stability coefficients are therefore not optional add-ons but essential components of responsible reporting.

Measurement considerations also warrant attention. Basic Gaussian Graphical Models treat observed variables as measured without error. Neuropsychological tests, however, vary in reliability and may share method variance. Observed associations may partly reflect measurement artefacts rather than cognitive interdependence. Latent network models and hybrid approaches can address some of these concerns, but they impose additional data requirements and modelling complexity.

Centrality metrics deserve particular caution. Measures such as strength, betweenness, and closeness were developed in social network contexts and do not automatically translate into clinical intervention targets. A highly central node may reflect measurement properties, scaling effects, or redundancy rather than functional importance. Interpretation must therefore remain theory-driven rather than metric-driven.

An additional consideration concerns data and node selection. As discussed in How to Create a Neuropsychological Network: A Step-by-Step Guide section, the interpretation of any neuropsychological network depends strongly on which variables are included in the model. This issue is particularly relevant when considering demographic variables, global cognitive screeners, or multiple tests from the same cognitive domain. Including demographic variables directly in the network estimates relations among cognitive tests conditional on those demographic influences; therefore, raw cognitive scores should generally be preferred in this case to avoid overcorrection. Conversely, when groups are already matched on demographic characteristics or when the study focuses on a restricted and homogeneous sample, demographic variables may be omitted. Similarly, including several highly overlapping measures from the same domain can redistribute shared variance and weaken individual edges. Thus, node selection should always be aligned with the research question, the study design, and the theoretical interpretation intended for the network.

Finally, researchers must remain aware of the distinction between between-person and within-person structure. Cross-sectional networks describe how individuals differ from one another; they do not necessarily describe how cognitive processes interact within a single individual over time. Conflating these levels can lead to misleading conclusions about mechanisms.

In sum, network psychometrics does not replace traditional neuropsychological models. Instead, it provides a complementary structural perspective. When applied with methodological rigour and theoretical restraint, it can illuminate how cognitive systems are organized, how they reorganize in pathology, and how interdependencies among tests contribute to clinical interpretation. When applied uncritically, however, it risks overinterpreting unstable associations or conflating description with explanation. The promise of network analysis in neuropsychology lies not in the production of visually compelling graphs, but in the disciplined examination of cognitive structure.

GLOSSARY

Core Concepts

Network (psychometric network)	A statistical representation of relationships among observed variables, in which nodes represent variables (e.g. test scores) and edges represent conditional associations between them.
Conditional dependence/conditional association	A statistical relationship between two variables that remains after controlling for all other variables in the network.
Conditional independence	The absence of an edge between two nodes, indicating that their association can be explained by other variables in the network.
Directed network	A network in which edges have a direction, indicating asymmetric relationships between variables.

Edge	A connection between two nodes representing their statistical association, typically after controlling for all other variables in the network.
Node	An individual variable in the network. In the context of Network Neuropsychology, it typically corresponds to a neuropsychological test score or measure.
Pairwise Markov Random Field (PMRF)	A class of statistical models in which variables are represented as nodes connected by edges encoding conditional dependencies; commonly used to estimate psychometric networks.
Undirected network	A network in which edges have no direction, indicating symmetric relationships between variables.
Unweighted network	A network where edges merely indicate the presence or absence of a connection between two nodes, without assigning numerical values to represent the strength or magnitude of that relationship.
Weighted network	A network in which edges have numerical values indicating the strength (and possibly direction) of associations between nodes.
Statistical Framework	
Covariance matrix	A matrix describing the pairwise covariances between variables.
EBIC (Extended Bayesian Information Criterion)	A model selection criterion that penalizes model complexity, commonly used to select the optimal level of regularization in network estimation.
Edge selection/pruning	The process by which weak or non-reliable edges are removed from the network, typically through regularization.
Gaussian Graphical Model (GGM)	A PMRF for continuous data in which edges correspond to partial correlations between variables.
Hyperparameter (γ, gamma)	A parameter in EBIC controlling the strength of the penalty for model complexity; higher values favour simpler models.
Inverse covariance matrix (precision matrix)	The inverse of the covariance matrix; its non-zero elements correspond to conditional dependencies between variables.
LASSO (Least Absolute Shrinkage and Selection Operator)	A regularization method that shrinks small edge weights toward zero, effectively removing weak or spurious associations.
Model selection	The process of choosing the best-fitting model among alternatives, balancing goodness-of-fit and model complexity.
Overfitting	A modelling problem in which the estimated network captures random noise rather than the true underlying structure, reducing generalizability.
Partial correlation	The association between two variables after removing the influence of all other variables in the dataset.

Regularization	A statistical technique that imposes penalties on model complexity to prevent overfitting and produce sparse, interpretable networks.
Sparsity	The property of a network having relatively few edges, reflecting a preference for simpler models.
Tuning parameter (λ, lambda)	A parameter controlling the strength of regularization; higher values lead to sparser networks.

Network Structure and Interpretation

Absence of edge	Indicates conditional independence between two variables given all others; it does not imply the absence of any relationship in a broader sense.
Betweenness	A centrality metric reflecting how often a node lies on the shortest path between other nodes; generally unstable and not recommended for interpretation in cross-sectional networks.
Bridge centrality	A measure of how strongly a node connects different communities or clusters within the network.
Closeness	A centrality metric based on the average distance of a node to all others; often unstable and difficult to interpret in psychometric networks.
Collider	A variable that is positively influenced by at least two distinct variables (A→C←B). Controlling for a collider forces an artificial trade-off (often appearing as an unexpected negative edge) between its underlying causes, a statistical phenomenon known as collider bias or Berkson's paradox.
Community (EGA)	A group of nodes within a network that are more strongly connected to each other than to nodes outside the group, typically identified using community detection algorithms.
Clustering coefficient	A centrality metric that quantifies the degree to which a node's neighbours are connected to each other, reflecting the overall tendency of the network to form tightly interconnected groups of nodes.
Degree	The number of edges connected to a node (primarily used in unweighted networks).
Differential variability	The degree to which differences in node variances (standard deviations) dictate node centrality. In network psychometrics, this is evaluated by correlating nodes' standard deviations with their centrality strengths. A high correlation indicates that the network structure may be artificially driven by statistical range restrictions rather than meaningful psychological relationships.
Edge sign	The direction of association between two variables, typically positive or negative.
Edge weight	The numerical value assigned to an edge, indicating the strength of association between two nodes.
Expected influence	A centrality measure that sums edge weights while preserving their sign, allowing differentiation between positive and negative connectivity.
Mean edge weight (global strength)	The average strength of all edges in the network, reflecting overall connectivity.

Network density	A measure of how many edges are present relative to the total number of possible edges.
Node strength	The sum of the absolute weights of all edges connected to a node; the most commonly used centrality measure in psychometric networks.
Predictability	The proportion of variance in a node explained by its neighbouring nodes, reflecting how well a variable can be predicted from the network.
Accuracy and Stability	
Bootstrap (non-parametric or case-dropping bootstrap)	A resampling method used to assess the stability and accuracy of network estimates by repeatedly sampling from the data.
Correlation-stability (CS) coefficient	An index quantifying the robustness of centrality measures under case-dropping resampling.
Edge-weight accuracy	The precision of estimated edge weights, typically evaluated using bootstrap confidence intervals.
Conceptual Safeguards	
Association versus causation	Network edges represent statistical associations, not causal relationships, particularly in cross-sectional data.
Between-person versus within-person	Cross-sectional networks reflect differences between individuals and do not necessarily represent processes occurring within individuals over time
Communities versus latent constructs	In Exploratory Graph Analysis (EGA), communities are interpreted as emergent dimensions within the data; however, they represent statistical groupings and should not be assumed to correspond directly to underlying cognitive constructs without theoretical validation.

AUTHOR CONTRIBUTIONS

Giorgia Tosi: Conceptualization; methodology; software; data curation; formal analysis; writing – original draft; writing – review and editing; visualization. **Daniele Romano:** Conceptualization; methodology; supervision; formal analysis; writing – original draft; writing – review and editing.

ACKNOWLEDGEMENT

Open access publishing facilitated by Universita degli Studi di Milano-Bicocca, as part of the Wiley - CRUI-CARE agreement.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data and code used in this tutorial are openly available at OSF repository (https://osf.io/73y5r/overview?view_only=35ff46d85a904506bf98b364b833f68f). The repository includes the dataset, annotated R Markdown script and all materials necessary to reproduce the analyses and figures presented in this article.

ORCID

Giorgia Tosi https://orcid.org/0000-0001-6405-3310
Daniele Romano https://orcid.org/0000-0002-6366-8433

REFERENCES

Altenbuchinger, M., Weihs, A., Quackenbush, J., Grabe, H. J., & Zacharias, H. U. (2020). Gaussian and mixed graphical models as (multi-)omics data analysis tools. *Biochimica et Biophysica Acta, Gene Regulatory Mechanisms*, *1863*(6), 194418. https://doi.org/10.1016/j.bbagrm.2019.194418

Ávalos-Tejeda, M., & Calderón, C. (2025). Psychometric network model recovery: The effect of sample size, number of items, and number of nodes. *European Journal of Investigation in Health, Psychology and Education*, *15*(11), 235. https://doi.org/10.3390/ejihpe15110235

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, *66*(3), 411–421. https://doi.org/10.1097/01.psy.0000127692.23278.a9

Blanchard, M. A., Contreras, A., Kalkan, R. B., & Heeren, A. (2023). Auditing the research practices and statistical analyses of the group-level temporal network approach to psychological constructs: A systematic scoping review. *Behavior Research Methods*, *55*(2), 767–787. https://doi.org/10.3758/s13428-022-01839-y

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A.-M., Wysocki, A. C., van Borkulo, C. D., van Bork, R., & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, *1*(1), 58. https://doi.org/10.1038/s43586-021-00055-w

Briganti, G., Scutari, M., Epskamp, S., Borsboom, D., Hoekstra, R. H. A., Golino, H. F., Christensen, A. P., Morvan, Y., Ebrahimi, O. V., Costantini, G., Heeren, A., de Ron, J., Bringmann, L. F., Huth, K., Haslbeck, J. M. B., Isvoranu, A.-M., Marsman, M., Blanken, T., Gilbert, A., … McNally, R. J. (2024). Network analysis: An overview for mental health research. *International Journal of Methods in Psychiatric Research*, *33*(4), e2034. https://doi.org/10.1002/mpr.2034

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., Wigman, J. T. W., & Snippe, E. (2019). What do centrality measures measure in psychological networks? *Journal of Abnormal Psychology*, *128*(8), 892–903. https://doi.org/10.1037/abn0000446

Burger, J., Isvoranu, A.-M., Lunansky, G., Haslbeck, J. M. B., Epskamp, S., Hoekstra, R. H. A., Fried, E. I., Borsboom, D., & Blanken, T. F. (2023). Reporting standards for psychological network analyses in cross-sectional data. *Psychological Methods*, *28*(4), 806–824. https://doi.org/10.1037/met0000471

Caffarra, P., Gardini, S., Zonato, F., Concari, L., Dieci, F., Copelli, S., Freedman, M., Stracciari, A., & Venneri, A. (2011). Italian norms for the Freedman version of the clock drawing test. *Journal of Clinical and Experimental Neuropsychology*, *33*(9), 982–988. https://doi.org/10.1080/13803395.2011.589373

Caltagirone, C., Gainotti, G., Carlesimo, G. A., & Parnetti, L. (1995). Batteria per la valutazione del deterioramento mentale: I. Descrizione di uno strumento di diagnosi neuropsicologica [The mental deterioration battery: I. Description of a neuropsychological diagnostic instrument]. *Archivio di Psicologia, Neurologia e Psichiatria*, *56*(4), 461–470.

Capasso, R., & Miceli, G. (2001). Esame Neuropsicologico per l'Afasia: E.N.P.A. In *Metodologie Riabilitative in Logopedia*. Springer. https://books.google.it/books?hl=it&lr=&id=4tzVIzJi6uwC&oi=fnd&pg=PA1&dq=Capasso+R.,+Miceli+G.+(2001).+Esame+Neuropsicologico+dell'afasia.+Milano.+Berlin:+Springer+Verlag.&ots=Jd-qBafxiT&sig=RbTyn76f8JoLODRc9zTOeJe0TxE&redir_esc=y#v=onepage&q&f=false

Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*(3), 759–771. https://doi.org/10.1093/biomet/asn034

Constantin, M. A. (2019). Sample size recommendations for estimating cross-sectional network models, *53*(9), 1689–1699.

Constantin, M. A., Schuurman, N. K., & Vermunt, J. K. (2026). A general Monte Carlo method for sample size analysis in the context of network models. *Psychological Methods*, *31*(3), 385–405. https://doi.org/10.1037/met0000555

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, R Special Issue*, *54*, 13–29. https://doi.org/10.1016/j.jrp.2014.07.003

Danaher, P., Wang, P., & Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *76*(2), 373–397. https://doi.org/10.1111/rssb.12033

De Marco, M., Wright, L. M., Bermejo, J. M. V., & Ferguson, C. E. (2024). APOE ε4 positivity predicts centrality of episodic memory nodes in patients with mild cognitive impairment: A cohort-based, graph theory-informed study of cognitive networks. *Neuropsychologia*, *192*, 108741.

Dubois, B., Andrade, K., & Levy, R. (2008). Executive dysfunction and neuropsychological testing. *Handbook of Clinical Neurology*, *89*, 35–52. https://doi.org/10.1016/S0072-9752(07)01203-1

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, *7*(1), 1–26. https://doi.org/10.1214/aos/1176344552

Epskamp, S. (2016). *Regularized Gaussian Psychological Networks: Brief Report on the Performance of Extended BIC Model Selection*. http://arxiv.org/abs/1606.05771

Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika*, *85*(1), 206–231. https://doi.org/10.1007/s11336-020-09697-3

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

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*(4), 1–18.

Epskamp, S., Isvoranu, A.-M., & Cheung, M. W.-L. (2022). Meta-analytic gaussian network aggregation. *Psychometrika*, *87*(1), 12–46. https://doi.org/10.1007/s11336-021-09764-3

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

Falkenström, F., Solomonov, N., & Rubel, J. (2020). Using time-lagged panel data analysis to study mechanisms of change in psychotherapy research: Methodological recommendations. *Counselling and Psychotherapy Research*, *20*(3), 435–441. https://doi.org/10.1002/capr.12293

Ferguson, C. E. (2021). A network psychometric approach to neurocognition in early Alzheimers's disease. *Cortex*, *137*, 61–73. https://doi.org/10.1016/j.cortex.2021.01.002

Ferguson, C. E. (2022). Network neuropsychology: The map and the territory. *Neuroscience & Biobehavioral Reviews*, *132*, 638–647. https://doi.org/10.1016/j.neubiorev.2021.11.024

Ferguson, C. E., & Foley, J. A. (2024). The influence of working memory and processing speed on other aspects of cognitive functioning in de novo Parkinson's disease: Initial findings from network modelling and graph theory. *Journal of Neuropsychology*, *18*(1), 136–153. https://doi.org/10.1111/jnp.12333

Folstein, M., Folstein, S., & McHugh, P. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198. https://doi.org/10.1002/(SICI)1099-1166(199805)13:5%253C285::AID-GPS753%253E3.0.CO;2-V

Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010*.

Freedman, M., Leach, L., Kaplan, E., Winocur, G., Shulman, K. I., & Delis, D. C. (1994). *Clock drawing: A neuropsychological analysis. Oxford University Press*.

Fried, E. I. (2020). Lack of theory building and testing impedes Progress in the factor and network literature. *Psychological Inquiry*, *31*(4), 271–288. https://doi.org/10.1080/1047840X.2020.1853461

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*(11), 1129–1164. https://doi.org/10.1002/spe.4380211102

Golino, H., Christensen, A. P., Moulder, R., Kim, S., & Boker, S. M. (2022). Modeling latent topics in social media using dynamic exploratory graph analysis: The case of the right-wing and left-wing trolls in the 2016 US elections. *Psychometrika*, *87*(1), 156–187. https://doi.org/10.1007/s11336-021-09820-y

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS One*, *12*(6), 1–26. https://doi.org/10.1371/journal.pone.0174035

Haslbeck, J. M. B., & Waldorp, L. J. (2015). Structure estimation for mixed graphical models in high-dimensional data (Preprint arXiv:1510.06871).

Haslbeck, J. M., & Waldorp, L. J. (2020). mgm: Estimating time-varying mixed graphical models in high-dimensional data. *Journal of Statistical Software*, *93*, 1–46.

Hevey, D. (2018). Network analysis: A brief overview and tutorial. *Health Psychology and Behavioral Medicine*, *6*(1), 301–328. https://doi.org/10.1080/21642850.2018.1521283

Huth, K. B. S., de Ron, J., Goudriaan, A. E., Luigjes, J., Mohammadi, R., van Holst, R. J., Wagenmakers, E.-J., & Marsman, M. (2023). Bayesian analysis of cross-sectional networks: A tutorial in R and JASP. *Advances in Methods and Practices in Psychological Science*, *6*(4), 25152459231193334. https://doi.org/10.1177/25152459231193334

JASP Team. (2025). *JASP [Computer software]*. (Versione Version 0.95.3) [Software].

Jones, P. J., Ma, R., & McNally, R. J. (2021). Bridge centrality: A network approach to understanding comorbidity. *Multivariate Behavioral Research*, *56*(2), 353–367. https://doi.org/10.1080/00273171.2019.1614898

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston naming test*. https://osf.io/vy8gh/download

Lauritzen, S. L. (1996). *Graphical models* (Vol. *17*). Oxford University Press. https://www.scribd.com/document/502433822/Lauritzen-Graphical-Models-1996

Liu, H., Han, F., Yuan, M., Lafferty, J., & Wasserman, L. (2012). The nonparanormal SKEPTIC. *arXiv*. 1206.6488. https://doi.org/10.48550/arXiv.1206.6488

Longo, C., Romano, D. L., Malaguti, M. C., Bacchin, R., & Papagno, C. (2024). Cognitive reorganization in patients with Parkinson's disease and mild cognitive impairment: A neuropsychological network approach. *Scientific Reports*, *14*(1), 28482.

Monaco, M., Costa, A., Caltagirone, C., & Carlesimo, G. A. (2013). Forward and backward span for verbal and visuo-spatial data: Standardization and normative data from an Italian adult population. *Neurological Sciences*, *34*(5), 749–754. https://doi.org/10.1007/s10072-012-1130-x

R Core Team. (2021). *R: A language and environment for statistical computing*. [Software]. R Foundation for Statistical Computing. https://www.R-project.org/

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

Romano, D., Tosi, G., & Alzheimer's Disease Neuroimaging Initiative. (2026). Divergent network trajectories in mild cognitive impairment: Converters versus non-converters to Alzheimer's disease. *Journal of Alzheimer's Disease*, *110*(3), 1414–1427.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Scharfenberg, D., Kalbe, E., Balzer-Geldsetzer, M., Berg, D., Hilker-Roggendorf, R., Kassubek, J., Liepelt-Scarfone, I., Mollenhauer, B., Reetz, K., Riedel, O., Roeske, S., Schulz, J. B., Storch, A., Trenkwalder, C., Witt, K., Wittchen, H. U., Dodel, R., & Ophey, A. (2025). A network perspective on cognition in individuals with Parkinson's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *17*(1), e70091.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society. Series A, Statistics in Society*, *73*(3), 273–282. https://doi.org/10.1111/j.1439-0264.1997.tb00113.x

Tosi, G., Borsani, C., Castiglioni, S., Daini, R., Franceschi, M., & Romano, D. (2020). Complexity in neuropsychological assessments of cognitive impairment: A network analysis approach. *Cortex*, *124*(December), 85–96. https://doi.org/10.1016/j.cortex.2019.11.004

Tosi, G., Nigro, S., Urso, D., Spinosa, V., Gnoni, V., Filardi, M., Giaquinto, F., Rizzi, E., Iaia, M., Macchitella, L., Chiarello, Y., Ferrari, F., Angelelli, P., Romano, D., & Logroscino, G. (2024). The network structure of cognitive impairment: From subjective cognitive decline to Alzheimer's disease. *Journal of Neuroscience*, *44*(27), e1344232023. https://doi.org/10.1523/JNEUROSCI.1344-23.2023

van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports*, *4*(1), 1–10. https://doi.org/10.1038/srep05918

van Borkulo, C. D., van Bork, R., Boschloo, L., Kossakowski, J. J., Tio, P., Schoevers, R. A., Borsboom, D., & Waldorp, L. J. (2023). Comparing network structures on three aspects: A permutation test. *Psychological Methods*, *28*(6), 1273–1285. https://doi.org/10.1037/met0000476

Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842

Warrington, E. K., & James, M. (1991). *The visual object and space perception battery. Thames Valley Test Company.*

Wright, L. M., De Marco, M., & Venneri, A. (2021). A graph theory approach to clarifying aging and disease related changes in cognitive networks. *Frontiers in Aging Neuroscience*, *13*, 676618.

Wysocki, A., van Bork, R., Cramer, A., & Rhemtulla, M. (2022). *Cross-lagged network models*. OSF. https://doi.org/10.31234/osf.io/vjr8z

Zoccolotti, P., Angelelli, P., Marinelli, C. V., & Romano, D. L. (2021). A network analysis of the relationship among reading, spelling and maths skills. *Brain Sciences*, *11*(5), 656.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Data S1. Supporting Information.

How to cite this article: Tosi, G., & Romano, D. (2026). Estimating and interpreting psychometric networks in neuropsychology: A tutorial with conceptual extensions. *Journal of Neuropsychology*, *00*, 1–27. https://doi.org/10.1111/jnp.70076