

**Relating teenagers' science interest network characteristics to later science course enrolment: An analysis of Australian PISA 2006 and LSAY data - Supplementary Materials**

### **Topological analysis**

Before estimating the networks, we performed a statistical analysis of the topological overlap of nodes using the *goldbricker* function of the *networktools* package in R (Jones, 2019) with the default settings ( $p = 0.05$ , `method = "hittner2003"`, `threshold = 0.25`, `corMin = 0.5`). If two nodes are very similarly related with all other nodes, they may redundantly measure the same underlying construct. If nodes are indeed too similarly related to the other nodes in the network, the better (i.e., more unique) node is retained.

The topological overlap analysis found that there was no significant topological overlap in any of the three SINMs.

### **Missing values**

In total, 678 missing values were identified in the data set. We investigated whether the missing values had a relationship with any values, observed or missing, by testing if it was missing completely at random. This was done using the *TestMCARNormality* function from the R package *MissMech* (Jamshidian, Jalal, & Jansen, 2014), which not only tests whether the missing data is missing completely at random (MCAR), but also whether the data have a multivariate normal distribution. Groups with similar patterns of missing data are identified and their equality of covariances are then tested (Jamshidian & Jalal, 2010).

We performed this analysis on the complete data set, including all four interest items. Four different patterns of missing data were found: 4180 cases had no missing data, 432 cases had missing data on self-concept and future motivation items, 8 cases had missing data on behavior items, 7 cases had missing data on one future motivation item (Fmw). The Hawkins test of normality and homoscedasticity was significant ( $p < .001$ ), meaning that assumptions of multivariate normality or homoscedasticity (or both) are not met in the data set. The non-parametric test was not significant ( $p = .234$ ), meaning that MCAR can be assumed, given that the data is not normally distributed.

As the missing values can be assumed to be MCAR, we decided to remove missing data case-wise, resulting in a final sample of N = 4180.

### Forecast analysis - model fit

To test whether centrality indices were predictors of the forecast correlation even when including the covariates node mean and node variance, we ran a covariance analysis including the covariates next to the centrality index. The linear regression model fit was then tested using the *stepAIC* function from the R package *MASS* (Venables & Ripley, 2002), a stepwise algorithm evaluating model fit based on the Akaike information criterion (AIC). Table S1 displays the results of the regressions for both the original model, including node mean, variance and centrality, as well as the final model, to allow for their comparison.

**Table S1**

Results of the model fit test of the linear regressions predicting the forecast correlation per SINM and centrality index.

Model	Predictor	<i>B</i>	<i>SE B</i>	$\beta$	<i>t</i>	<i>p</i>
Chemistry - strength						
Original model	$F(3, 27) = 13.13, p < .001, R^2_{Adjusted} = .55, AIC = -147.70$					
	Mean	.14	.10	.19	1.47	.152
	Variance	.09	.41	.03	.23	.821
	Strength	.06	.01	.75	5.71	<.001
Final model	$F(2, 28) = 20.36, p < .001, R^2_{Adjusted} = .56, AIC = -149.64$					
	Mean	.15	.09	.21	1.70	.101
	Strength	.06	.01	.76	6.28	<.001
Chemistry - closeness						
Original model	$F(3, 27) = 12.20, p < .001, R^2_{Adjusted} = .53, AIC = -146.36$					

	Mean	.06	.10	.08	.57	.57
	Variance	1.22	.40	.41	3.06	.005
	Closeness	65.54	11.94	.70	5.49	<.001
Final model	$F(2, 28) = 18.58, p < .001, R^2_{Adjusted} = .54, AIC = -147.99$					
	Variance	1.29	.37	.43	3.44	.002
	Closeness	65.25	11.78	.70	5.54	<.001
<hr/>						
Physics - strength						
Original model	$F(3, 27) = 5.12, p = .006, R^2_{Adjusted} = .29, AIC = -135.28$					
	Mean	.07	.12	.10	.62	.54
	Variance	.57	.49	.20	1.26	.25
	Strength	.04	.02	.51	3.15	.004
Final model	$F(2, 28) = 7.66, p = .002, R^2_{Adjusted} = .31, AIC = -136.85$					
	Variance	.67	.46	.23	1.48	.15
	Strength	.04	.01	.49	3.13	.004
<hr/>						
Physics - closeness						
Original model	$F(3, 27) = 5.89, p = .003, R^2_{Adjusted} = .33, AIC = -136.93$					
	Mean	.02	.11	.03	.16	.87
	Variance	1.23	.46	.42	2.66	.01
	Strength	45.71	13.22	.52	3.46	.002
Final model	$F(2, 28) = 9.14, p < .001, R^2_{Adjusted} = .35, AIC = -138.90$					
	Variance	1.24	.43	.43	2.90	.007

	Closeness	45.70	12.99	.52	3.52	.002
<b>Biology - strength</b>						
Original model	$F(3, 28) = 4.03, p = .017, R^2_{Adjusted} = .23, AIC = -139.73$					
Mean	.30	.12	.43	2.55	.017	
Variance	-.08	.49	-.03	-.16	.87	
Strength	.03	.01	.40	2.38	.024	
Final model	$F(2, 29) = 6.25, p = .006, \text{adjusted } R^2 = .25, AIC = -141.70$					
Mean	.30	.11	.42	2.69	.012	
Strength	.03	.01	.39	32.53	.017	
<b>Biology - closeness</b>						
Original model	$F(3, 28) = 4.81, p = .008, R^2_{Adjusted} = .27, AIC = -141.53$					
Mean	.25	.11	.37	2.29	.030	
Variance	.49	.45	.17	1.07	.294	
Strength	37.11	13.41	.43	2.77	.010	
Final model	$F(2, 29) = 6.61, p = .004, R^2_{Adjusted} = .27, AIC = -142.25$					
Mean	.28	.11	.41	2.68	.012	
Closeness	35.31	13.34	.41	2.65	.013	

## Descriptive Statistics and covariance matrices

The descriptive statistics of all included interest components before dichotomizing can be found in Table S2. The interest items were taken together in the science interest component and not separately, as done in the SINMs.



**Table S2**

Means and standard deviations of the science interest components for the whole sample and per gender

Group	<i>N</i>	Interest	Enjoyment	Behavior	Value	Self-concept	Future Intentions	Knowledge
Total	4180	2.70 (.94)	2.75 (.79)	1.53 (.63)	2.87 (.80)	2.65 (.75)	2.21 (.94)	563.53 (90.29)
Boys	1893	2.71 (.94)	2.83 (.79)	1.60 (.67)	2.94 (.79)	2.77 (.74)	2.29 (.94)	570.20 (89.96)
Girls	2287	2.70 (.92)	2.68 (.79)	1.48 (.59)	2.82 (.80)	2.55 (.74)	2.14 (.94)	558.01 (89.82)

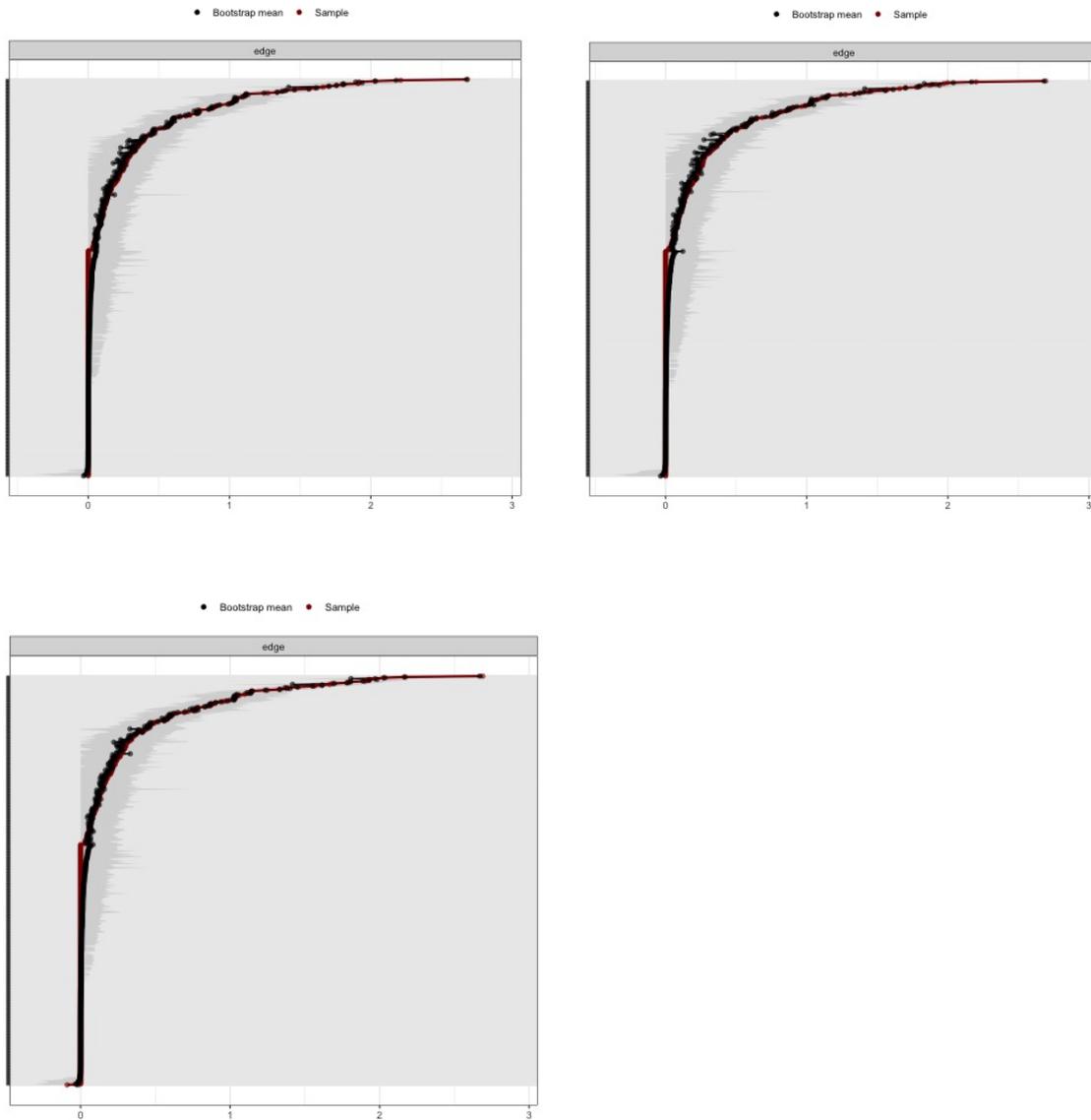
## Additional analyses

### 1. Network stability and accuracy

#### Accuracy and stability estimation

Assessment of the accuracy and stability of the three SINMs was done using functions included in the R- package *bootnet* (Epskamp, Borsboom & Fried, 2018). We applied the edge weight bootstrap as well as the subsetting bootstrap on the three SINMs. More information on these methods can be found in the tutorial paper on accuracy and stability of psychological networks and the use of the R-package *bootnet* by Epskamp and colleagues (2018).

**Edge weight stability.** First, in order to estimate the edge weight accuracy of the SINMs, 95% confidence intervals (CIs) of the edge weights were bootstrapped. 2000 nonparametric bootstraps were run per SINM, with *IsingFit* as default method. Visualizations of the edge weight bootstraps are displayed in Figure S1.



*Figure S1. Top left.* Edge weight stability plot for the chemistry course SINM. *Top right.* Edge weight stability plot for the physics course SINM. *Bottom left.* Edge weight stability plot for the biology course SINM. Edges connecting two nodes are represented by the points in the graph - with the black points representing the bootstrap mean and the red line indicates the estimated edge weights based on the sample. The grey area around the line represents the bootstrapped CIs per edge. Strength of the edge weights are represented on the x-axis and all possible edges are represented on the y-axis, ordered based on edge weight - from highest edge weight (top) to lowest edge weight (bottom).

**Centrality stability.** Second, we examined the centrality indices' stability by using the subsetting bootstrap method, in which the network and its centrality indices are re-estimated a number of times (1000 times in our analyses) with a decreasing number of participants. A correlation of the order of central indices of the original and the re-estimated network is then computed, so that a high correlation is an indicator for stability of the centrality indices. The resulting, so-called centrality stability coefficient (CS-coefficient), should be at least .25, but preferably above .5 for the indices to be seen as stable (Epskamp, et al., 2018); for results of the CS-coefficient, see the section titled Forecast analysis in the main article.

Centrality stability was estimated for three centrality indices: betweenness, closeness and strength. See Figure S2 for the results of the subsetting bootstraps of the three SINMs.

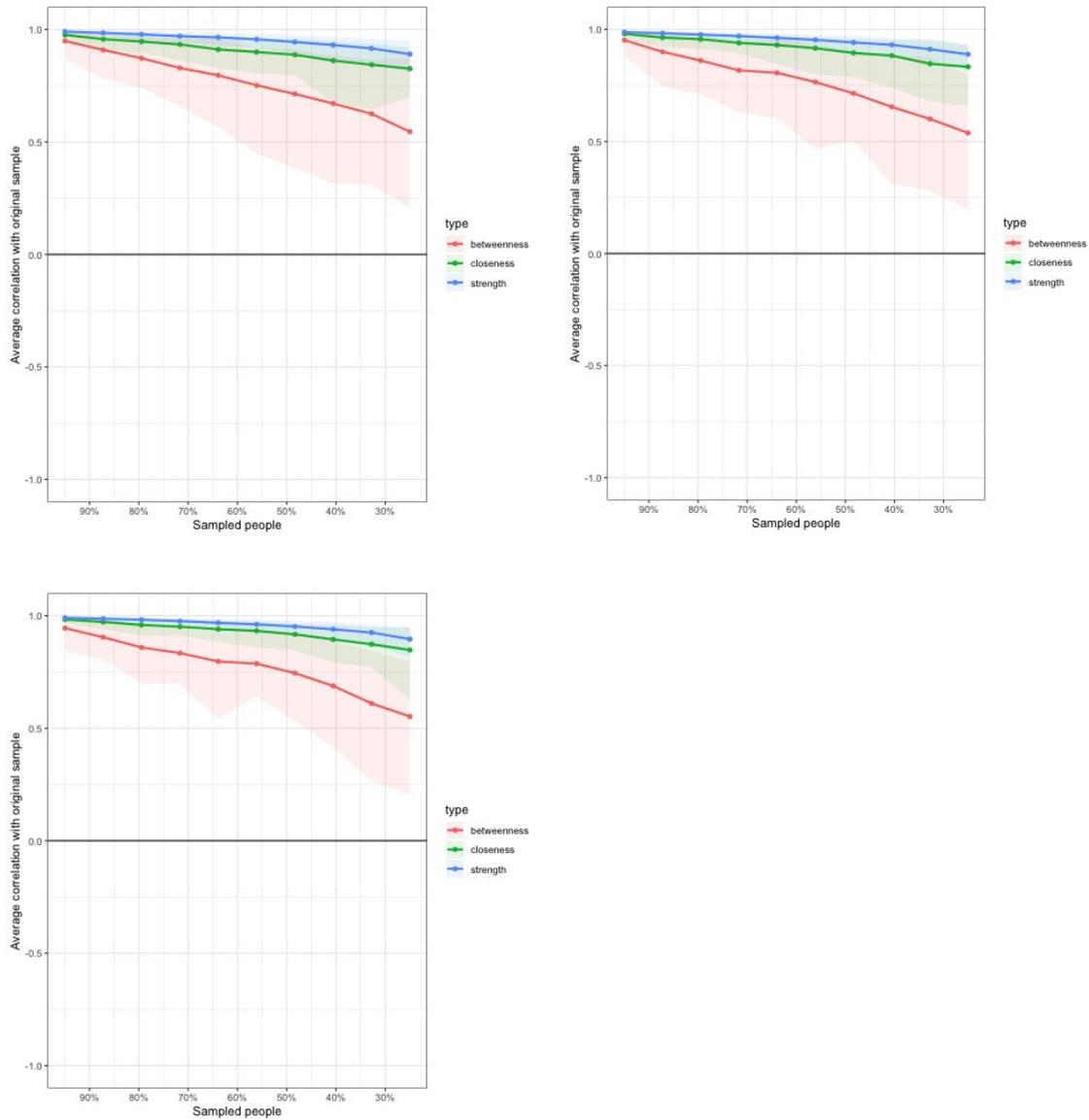


Figure S2. *Top left.* Subsetting bootstrap for the chemistry course SINM. *Top right.* Subsetting bootstrap for the physics course SINM. *Bottom left.* Subsetting bootstrap for the biology course SINM. The average correlation of the original network indices (full sample) and the indices estimated from the smaller subsamples of the data set is shown per centrality index. The x-axis represents the percentage of sampled cases and the y-axis represents the correlation between original and subsample. The lines represent the mean correlations per centrality index and the surrounding areas indicate the range from the 2.5th to the 97.5th quantile (Epskamp et al., 2018).

## 2. Difference tests

The bootstrapped difference test can be used to compare edge weights and centralities (Epskamp et al., 2018). The results of the non-parametric bootstraps are used to compare both the edge weights and the centralities; to obtain these results, we set the *statistics* argument to “edge”, “strength” and “closeness”. Betweenness was not included, as it was not stable. As we were not interested in specific edges, but in centrality indices, we did not investigate the edge weight bootstrapped difference test.

**Centrality bootstrapped difference tests.** We conducted centrality bootstrapped difference tests for the centrality indices per course SINM that were significantly correlated with the forecast correlation - resulting in four bootstrapped difference tests: Node strength in all the course SINMs and node closeness in the biology SINM. The bootstrapped difference tests, with an expected significance level of approximately  $p < .05$ , are displayed in Figure S3.



*Figure S3. Top left:* Difference test for node strength in the chemistry course SINM. *Top right:* Difference test for node closeness in the chemistry course SINM. *Middle left:* Difference test for node strength in the physics course SINM. *Middle right:* Difference test for node closeness in the physics course SINM. *Bottom left:* Difference test for node strength in the biology course SINM. *Bottom right:* Difference test for node closeness in the biology course SINM. Grey boxes represent nodes that do not significantly differ from each other on the given centrality index. Black boxes represent nodes that do significantly differ on the given centrality index. In the white boxes, the centrality of the value of the given centrality index is displayed.

### 3. NCT Comparison PV1 vs. PV2

To test whether using different knowledge PVs has an effect on the network structure and characteristics, we compared the networks using the network comparison test (NCT; van Borkulo, Epskamp, & Millner, 2016). The tested for network invariance on (1) network structure; (2) global connectivity and (3) centrality indices. The NCT uses permutations (i.e., repeated rearrangements of the samples) to test network invariance. We used 1000 permutation iterations and controlled for multiple comparison using Bonferroni correction.

*Chemistry.* (1) The networks using PV1 and PV2 did not differ significantly in network structure ( $M = .24, p = 1$ ); (2) global strength ( $S = 1.75, p = .395$ ) and (3) centrality indices, with all  $p = 1$ .

*Physics.* (1) The networks using PV1 and PV2 did not differ significantly in network structure ( $M = .25, p = 1$ ); (2) global strength ( $S = .48, p = .818$ ) and (3) centrality indices, with all  $p = 1$ .

*Biology.* (1) The networks using PV1 and PV2 did not differ significantly in network structure ( $M = .24, p = 1$ ); (2) global strength ( $S = .58, p = .776$ ) and (3) centrality indices, with all  $p = 1$ .

### 4. Forecast correlation based on different network estimation method

We computed the forecast correlation based on SINMs that were estimated using other estimation methods than *IsingFit*, which is used for binary data. All networks were estimated using the *estimateNetwork* function of the R package *bootnet* (Epskamp & Fried, 2017), using different defaults, which indicate the method to use for network estimation.

First we ran two different *ggmModSelect* models using the original variables, one using Pearson correlations (using `corMethod = "cor"`) and one using polychoric correlations (using `corMethod = "corAuto"`). Please note that when using Pearson correlations ("`cor`" method), all variables are considered being continuous, whereas the self-report measures are being considered as being ordinal when using the "`corAuto`" method, because they have less than seven unique integer values (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012). This type of network estimation can be used if a data set has a sufficiently large observations to node ratio as regularization is not needed (Williams & Rast, 2018) - which is the case in the data set we used. The unregularized Gaussian Graphical model (GGM) is estimated as follows (Epskamp & Fried, 2017): The model is chosen by first running the graphical least absolute shrinkage and selection operator (*glasso*) for 100 different tuning parameters to obtain 100 different network structures (from dense to sparse). In a second step, the best model is chosen according to the Bayesian information criterion (*BIC*) and lastly, all possible ways to improve the *BIC* are tested by adding and removing edges.

Next, we ran a *mgm* model, which can incorporate different types of measurement scales (continuous, categorical, binary) within the same model (Haslbeck & Waldorp, 2015). As the behaviour indicators were highly skewed, we included them as binary variables (1 = 0; 2-4 = 1) in the *mgm* model. Before running the forecast correlation per course, centrality and estimation method, we tested whether centrality indices were stable using the *bootnet* package in R (Epskamp & Fried, 2017). The centrality indices using the *ggmModSelect* "`cor`" method were as follows: strength ( $CS[cor=0.7] = .75, .75, .75$ ), closeness ( $CS[cor=0.7] = .59, .44, .44$ ) and betweenness centrality ( $CS[cor=0.7] = .36, .36, .36$ ) for the chemistry, physics and biology SINM, respectively. The centrality indices using the *ggmModSelect* "`cor_auto`" method were as follows: strength ( $CS[cor=0.7] = .44, .59, .75$ ), closeness ( $CS[cor=0.7] = .59, .59, .36$ ) and betweenness centrality ( $CS[cor=0.7] = .13, .13, .28$ ) for the chemistry, physics and biology SINM, respectively. The centrality indices using the *mgm* method were as follows: strength ( $CS[cor=0.7] = .75, .75, .75$ ), closeness ( $CS[cor=0.7] = .00, .00, .00$ ) and betweenness centrality ( $CS[cor=0.7] = .00, .00, .00$ ) for the chemistry, physics and biology SINM, respectively. Strength thus was sufficiently stable across all SINMs, irrespective of estimation methods, whereas the same was not true for closeness and betweenness.

The forecast correlations per course, centrality index and estimation method are displayed in Table S3. Note that the degrees of freedom differ per correlation as we used the topological analysis (Jones, 2019), which did indicate similar nodes when using non-binarized data.

**Table S3**

Forecast correlations per course, estimation method and centrality index.

	Strength centrality		Closeness centrality	
	<i>Pearson correlation</i>	<i>p-value</i>	<i>Pearson correlation</i>	<i>p-value</i>
Chemistry				
ggmModSelect "cor"	$r(29) = .52$	.003	$r(29) = .40$	.027
ggmModSelect "cor_auto"	$r(29) = .24$	.196	$r(29) = -.05$	.773
mgm	$r(29) = -.35$	.056	$r(29) = .16$	.402
Physics				
ggmModSelect "cor"	$r(28) = .39$	.034	$r(28) = .23$	.220
ggmModSelect "cor_auto"	$r(28) = .05$	.797	$r(28) = -.15$	.423
mgm	$r(28) = -.38$	.038	$r(28) = -.03$	.856
Biology				
ggmModSelect "cor"	$r(29) = .39$	.032	$r(29) = .29$	.118
ggmModSelect "cor_auto"	$r(29) = .05$	.795	$r(29) = -.09$	.647
mgm	$r(29) = -.29$	.120	$r(29) = -.08$	.666

The results were not stable across estimation methods: Strength and centrality based on the networks using *IsingFit* (original analysis) are both significantly positively correlated with the outcome measure across all three course SINMs. The centrality indices of the *ggmModSelect* SINMs using Pearson correlations (“cor”) are generally positively correlated, with strength being significantly positively correlated with the outcome for all three courses and closeness only being significantly correlated with the outcome for the chemistry SINM. For SINMs estimated with *ggmModSelect* using polychoric correlations (“cor\_auto”), centrality was not significantly correlated with the outcome. Lastly, only the strength centrality within the SINMs estimated with *mgm* are significantly correlated with the outcome in the physics SINM with strength having a negative correlation with the outcome.

Given (1) the difference in measurement levels across variables, with knowledge scores being continuous and self-reported measures being ordinal (measured on a 1-4 Likert scale), and (2) the finding that the assumptions of multivariate normality or homoscedasticity (or both) are not met for the data, we believe that estimating the SINMs using dichotomized data is the best option. Using *IsingFit* based on the dichotomized data, we are not bound by assumptions of multivariate normality or homoscedasticity, and the model is the simplest model of the four models we fit - being characterized by two parameters: a node threshold parameter and an edge connectivity parameter (van Borkulo & Epskamp, 2015). Lastly (3), centrality stability for strength and closeness was highest for the *IsingFit* method.

## References

- Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*, 195–212.  
<http://dx.doi.org/10.3758/s13428-017-0862-1>
- Epskamp, S., Cramer, A. O., 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., & Fried, E. I. (2017). bootnet: Bootstrap methods for various network estimation routines. <https://cran.r-project.org/web/packages/bootnet/index.html>
- Haslbeck, J., & Waldorp, L. J. (2015). mgm: Estimating time-varying mixed graphical models in high-dimensional data. *arXiv preprint arXiv:1510.06871*.
- Jamshidian, M., & Jalal, S. (2010). Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika*, *75*(4), 649-674.
- Jamshidian, M., Jalal, S. J., & Jansen, C. (2014). MissMech: An R package for testing homoscedasticity, multivariate normality, and missing completely at random (MCAR). *Journal of Statistical software*, *56*(6), 1-31.
- Jones, P. (2019). networktools: Tools for identifying important nodes in networks. R package version 1.2.3. <https://CRAN.R-project.org/package=networktools>
- van Borkulo, C. D., & Epskamp, S. (2015). IsingFit: Fitting Ising models using the eLasso method. R package version 0.3.0. Retrieved from <https://rdr.io/cran/IsingFit/>
- van Borkulo, C. D., Epskamp, S., & Millner, A. (2016). Network Comparison Test: Statistical comparison of two networks based on three invariance measures.  
<https://cran.r-project.org/web/packages/NetworkComparisonTest/index.html>
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* 4 edition Springer. *New York*.