Supplemental material for

Comparing Network Structures on Three Aspects: A Permutation Test

Claudia D. van Borkulo[1,2], Riet van Bork[3], Lynn Boschloo[4], Jolanda J. Kossakowski[5], Pia Tio[1], Robert A. Schoevers[6], Denny Borsboom[1], and Lourens J. Waldorp[1]

[1] Department of Psychological Methods, University of Amsterdam

[2] Centre for Urban Mental Health, University of Amsterdam

[3] Center for Philosophy of Science, University of Pittsburgh

[4] Department of Clinical, Neuro, and Developmental Psychology, VU University Amsterdam

[5] Research and Documentation Centre (WODC) of the Dutch Ministry of Security and Justice

[6] University of Groningen, University Medical Center Groningen, Dept of Psychiatry, Interdisciplinary Center Psychopathology and Emotion regulation

Appendix 1. Proof

Appendix 2. Figures

Appendix 1. Proof

In this appendix, we provide a proof for the validity of NCT for the tests on invariance of global strength, network structure, and edge strength. The proof is based on the idea that the resampling test is equivalent to the test based on the normal approximation, as shown by Van der Vaart (1998). In a resampling test, the critical value is estimated (or exact for small samples) and converges to the critical value of a corresponding parametric test. Consequently, only if a corresponding parametric test can be constructed for a hypothesis, an adequate resampling test can be constructed for it. The proof works for generalized linear models (i.e., exponential family models) and is, therefore, not limited to normally distributed random variables. The proof requires several assumptions, which are summarized at the end of the proof.

Proof. First, we show that the network parameters converge to a normal distribution (see assumption 1). Then, we show that a parametric test can be constructed and, hence, that a resampling test performs adequately.

We start with the linear regression model, which is at the basis of the network estimation methods in the accompanying paper. We set some X_i to be Y and then use regression on the remaining variables X_j for all $j \neq i$, which are collected in the matrix X. The total number of variables is p and $q = p - 1$. The regression is then

$$Y = X\omega_i + e,$$

where X is called the design matrix and has dimensions $n \times q$ (i.e., number of observations times the number of variables minus the variable that has been assigned Y), ω_i is the true coefficient corresponding to $X_i = Y$ and e is the error. X is assumed to be independent across the q variables and normally distributed with mean 0 and variance σ_e^2 (assumption

2).

Knight and Fu (2000) show that when the design matrix is non-singular (i.e., it is a full rank design matrix X in a linear regression model and $n > q$) and the lasso penalty λ grows at the same rate as $\sqrt{n}$ (i.e., $\lambda/\sqrt{n} \to \lambda_0$, which is finite; assumption 3), then the lasso estimator converges to a (biased) normal distribution. Tibshirani (1996) shows that in this full rank case we can approximate the lasso with the ridge estimator and hence the bias is approximately $(I - \Delta)\omega_i$, where Δ is defined as

$$\Delta = (X^\top X + \lambda W^-)^{-1} X^\top,$$

and W is diagonal with elements $|\hat{\omega}_{ij}|$, for $j \neq i$. If $\lambda = 0$, we obtain the ordinary least squares estimate. Tibshirani (1996) also uses the ridge estimator to obtain approximate standard errors. The covariance matrix using this approximation is given by

$$\Sigma = \hat{\sigma}^2 (X^\top X + \lambda W^-)^{-1} X^\top X (X^\top X + \lambda W^-)^{-1},$$

where $\hat{\sigma}^2$ is the estimate of the residual variance. The standard errors are obtained by taking the square root of the diagonal elements of Σ (note that the standard error is 0 when $\hat{\omega}_{ij}$ is 0; we do not mean for this approximation to be used in practice, we use it here merely for the proof). Consequently, using both results of Knight and Fu (2000), and Tibshirani (1996), and the assumption that for each variable the mean and variance are finite, we obtain that, approximately, the lasso estimator is distributed as

$$\hat{\omega_i} \sim N(\Delta\omega_i, \Sigma).$$

This shows that for linear regression we obtain an approximate normal distribution whenever $n > q$.

We can extend this result to generalized linear models (e.g., to logistic regression models for binary data) in the following way. Following Van de Geer et al. (2014), we can replace X from linear regression with the second-order derivative of the loss function L with respect to the parameters, denoted $\ddot{L}$ (a $q \times q$ matrix). By additionally assuming that this second-order derivative $\ddot{L}$ exists, we can plug-in $\ddot{L}$ for $X^\top X$ above and obtain a similar approximation for the normal distribution.

According to the resampling method, we repeatedly resample two data sets from the pooled data (without replacement), matching the original sample sizes. Estimates based on resampled data are indicated by an asterisk: $\hat{\omega}^{1*}$ and $\hat{\omega}^{2*}$.

For global strength, the test statistic is defined as

$$S(\hat{\omega}^{1*}, \hat{\omega}^{2*}) = |\sum_{i=1}^{p} \sum_{j>i} (|\hat{\omega}_{ij}^{1*}| - |\hat{\omega}_{ij}^{2*}|)|,$$

where $\hat{\omega}^{1*}$ is the vector with $p(p-1)/2$ unique edge estimates of network 1, and the sum is taken of all $p(p-1)/2$ edges in the network. Since $\hat{\omega}$ (using the ridge approximation) is a linear function of the data Y, we can immediately apply Theorem 13.25 of van der Vaart (1998). This requires (a) the the variance to be finite, and (b) n_1/N to converge to some constant in $(0, 1)$ as n_1 and n_2 become infinitely large (assumption 4). The first requirement is reasonable when variables are used whose distribution belong to the exponential family. The second implies that we cannot have group sizes that are extremely different. Given these assumptions, the critical value of the resampling statistic converges to the one of the normal distribution used for the parametric test. As a corollary, we immediately obtain that the edge test also has a parametric counterpart to the resampling test and, hence, is also valid.

For the test on invariant network structure, we use the maximum of all differences between corresponding estimates of both networks, i.e.,

$$M(\hat{\omega}^1, \hat{\omega}^2) = \max_{\substack{1 \leq i \leq p \\ j < i}} |\hat{\omega}^1_{ij} - \hat{\omega}^2_{ij}|.$$

For the resampling test to work, we use the fact that this is an extreme value statistic. Accordingly, the generalized extreme value theorem for standardized random variables $z = (x - \mu)/\sigma$, with finite mean μ and variance σ^2, shows that the distribution of the maximum statistic converges to one of three possible distributions: the Gumbel, Fréchet or Weibull distribution (see, e.g., Coles et al., 2001). Note that convergence in distribution is defined here in terms of the size of the network, so that the approximation is better when the network is larger.

Additionally, because the edges ($\hat{\omega}_{ij}$) are dependent, we require the assumption that the dependence is limited in the sense that some edges can be correlated but the correlation should approach zero for others (i.e., the correlation ρ_i tends to 0 at least at the rate of $1/\log i$; DasGupta, 2008, Theorem 8.16; assumption 5).

If the dependence of the estimates is limited in the correlation among the edge estimates, then we can apply Theorem 5.1 from Coles et al. (2001), where the maximum of the estimates will converge to the Gumbel distribution (in this case because the distribution of the estimates is approximately normal). For the Gumbel distribution we can obtain the critical value and hence for the resampling statistic we obtain a critical value that converges to the parametric one of the Gumbel distribution (see, e.g., DasGupta, 2008, Theorem 31.2 for a general result on this).

To conclude, there are five assumptions for NCT. Four assumptions are required for all tests (i.e., global strength, individual edge strength, and network structure invariance). Additionally, a fifth assumption is required for the test on network structure invariance. The assumptions for all three test are

1. The matrix $X^\top X$ or the second-order derivative of the loss function $\ddot{L}$ is non-singular (i.e., no multi-collinearity and more observations then variables).

2. The mean μ and variance σ^2 of each of the p variables are finite.

3. The lasso penalty λ grows at the same rate as $\sqrt{n}$, i.e., $\lambda/\sqrt{n} \to \lambda_0$ as $n \to \infty$ (i.e., λ can become larger with large samples but it cannot grow faster).

4. The group sizes n_1 and n_2 cannot be too unbalanced, i.e., n_1/N converges to a constant in $(0, 1)$, where $N = n_1 + n_2$.

In addition to these assumptions, the test on invariance of network structure requires the following.

5. The correlation function $\rho_i \log i \to 0$ as $i \to \infty$ (i.e., the edge strengths can be correlated to some extent but should go to zero for others).

□

References

Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*. London: Springer.

DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. Springer Science & Business Media.

Knight, K, & Fu, W. (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics*, *28*(5), 1356 – 1378.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*. Series B (Methodological), *58*, 267–288.

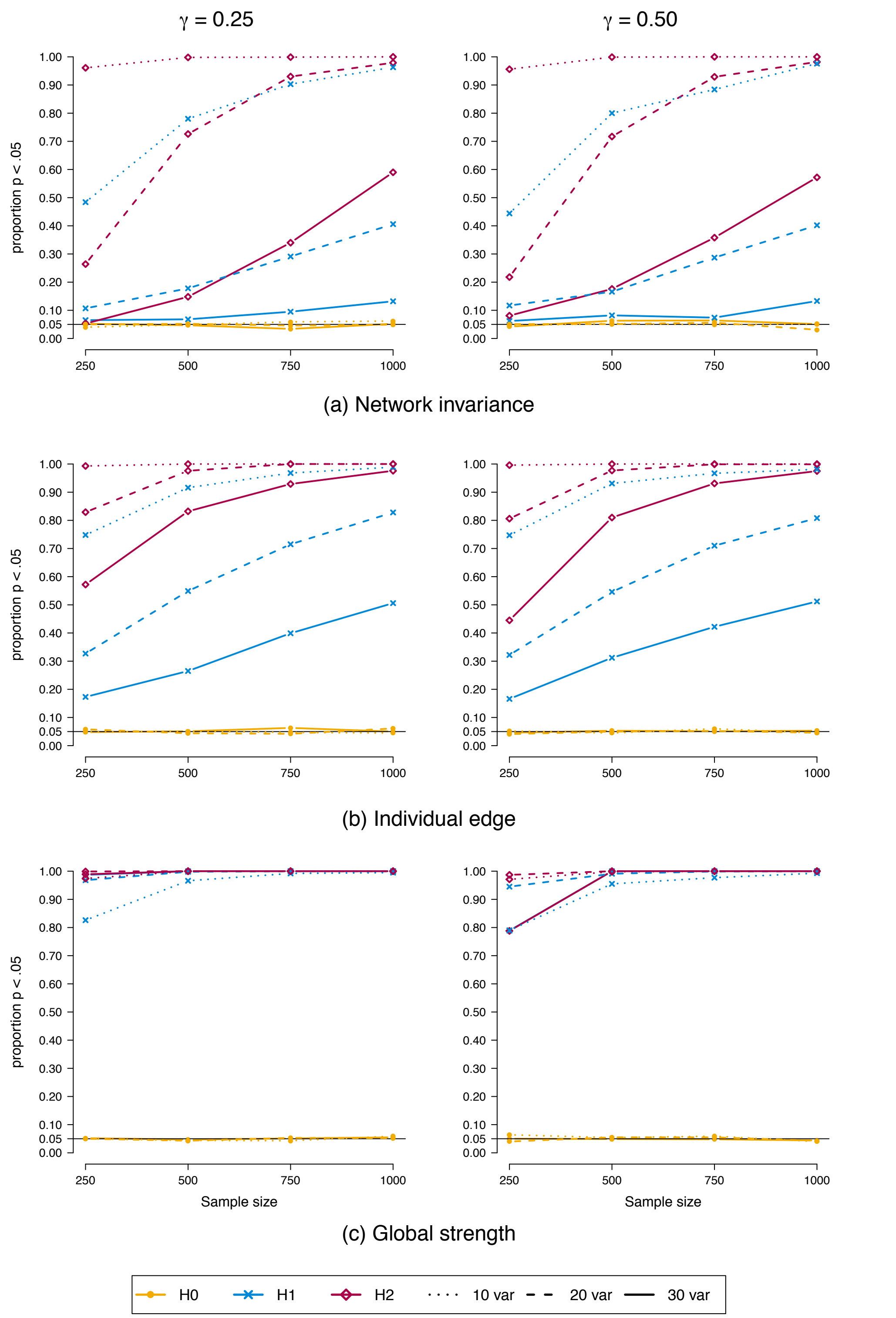
Van de Geer, S., Bühlmann, P., Ritov, Y. A., & Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3), 1166-1202.

Van der Vaart, A. W. (1998). *Asymptotic statistics.* New York: Cambridge University Press.

Appendix 2. Figures

Figure S1

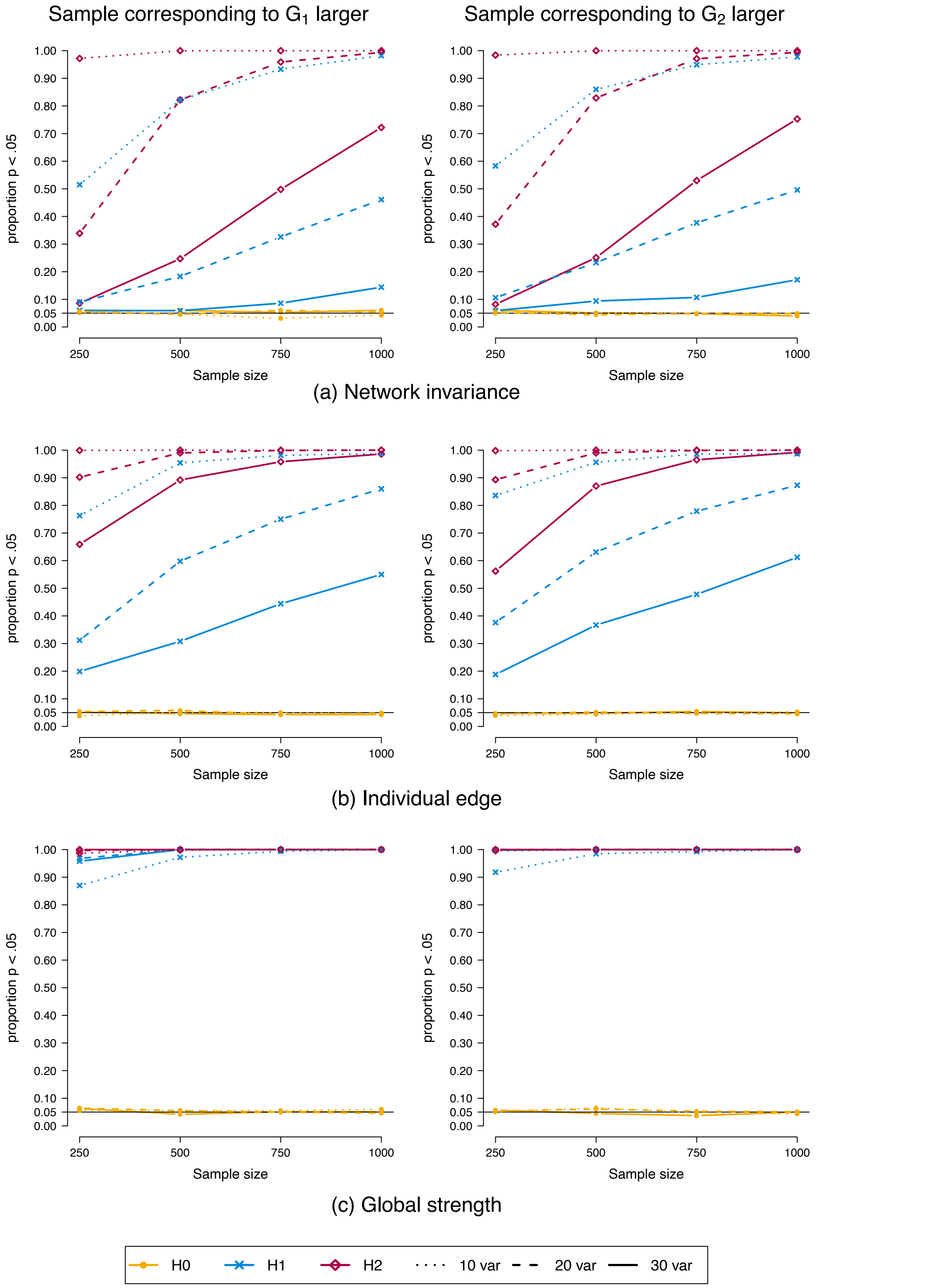
Performance of NCT with density = 0.3, γ = 0.25 and γ = 0.50 (main analysis)



Note. Simulation study results for tests on invariance of (a) network structure, (b) an individual edge, and (c) global strength test with continuous data (main analysis). The x-axes display sample size, whereas the y-axes display proportion of *p*-values < 0.05. All three tests were applied to simulated data under the null hypothesis of no difference (H_0; yellow) and under the alternative hypotheses that there is a difference to a certain degree (H_1, blue; and H_2, red). Data were simulated from networks with density 0.3 and different numbers of variables (10: dotted lines, 20: dashed lines, 30: solid lines). The data were simulated with equal sample sizes.

Figure S2

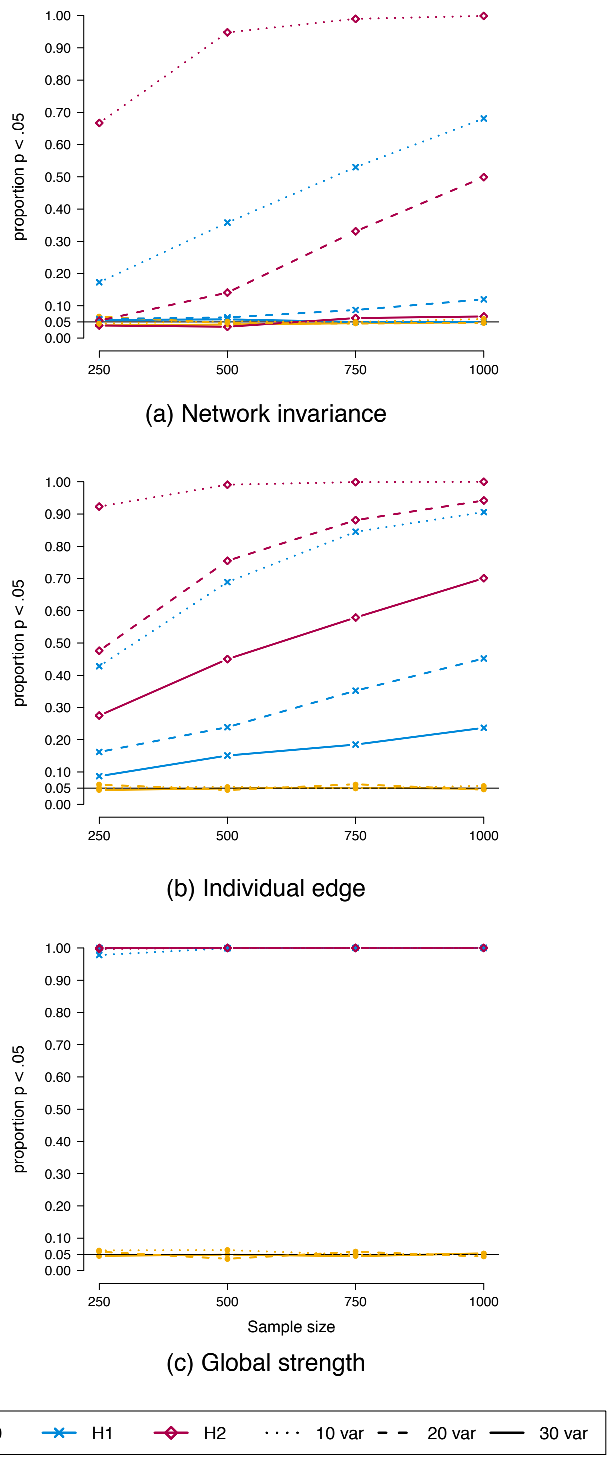
Performance of NCT with density = 0.3, γ = 0, and unequal sample sizes (main analysis)



Note. Simulation study results for tests on invariance of (a) network structure, (b) an individual edge, and (c) global strength test with continuous data (main analysis). All three tests were applied to simulated data under the null hypothesis of no difference (H_0, yellow) and under the alternative hypotheses that there is a difference to a certain degree (H_1, blue; and H_2, red). The x-axes display sample size, whereas the y-axes display proportion of *p*-values < 0.05. Data were simulated with unequal sample sizes: the largest sample was generated with either network G_1 (left panels) and were 1.5 times larger, or the largest sample was generated with network G_2 (right panels). Data were simulated from networks with different numbers of variables (10: dotted lines, 20: dashed lines, 30: solid lines). The networks all had density = 0.3 and the networks were all estimated with hyperparameter γ = 0.

Figure S3

Performance of NCT with density = 0.5 and γ = 0 (main analysis)



(a) Network invariance

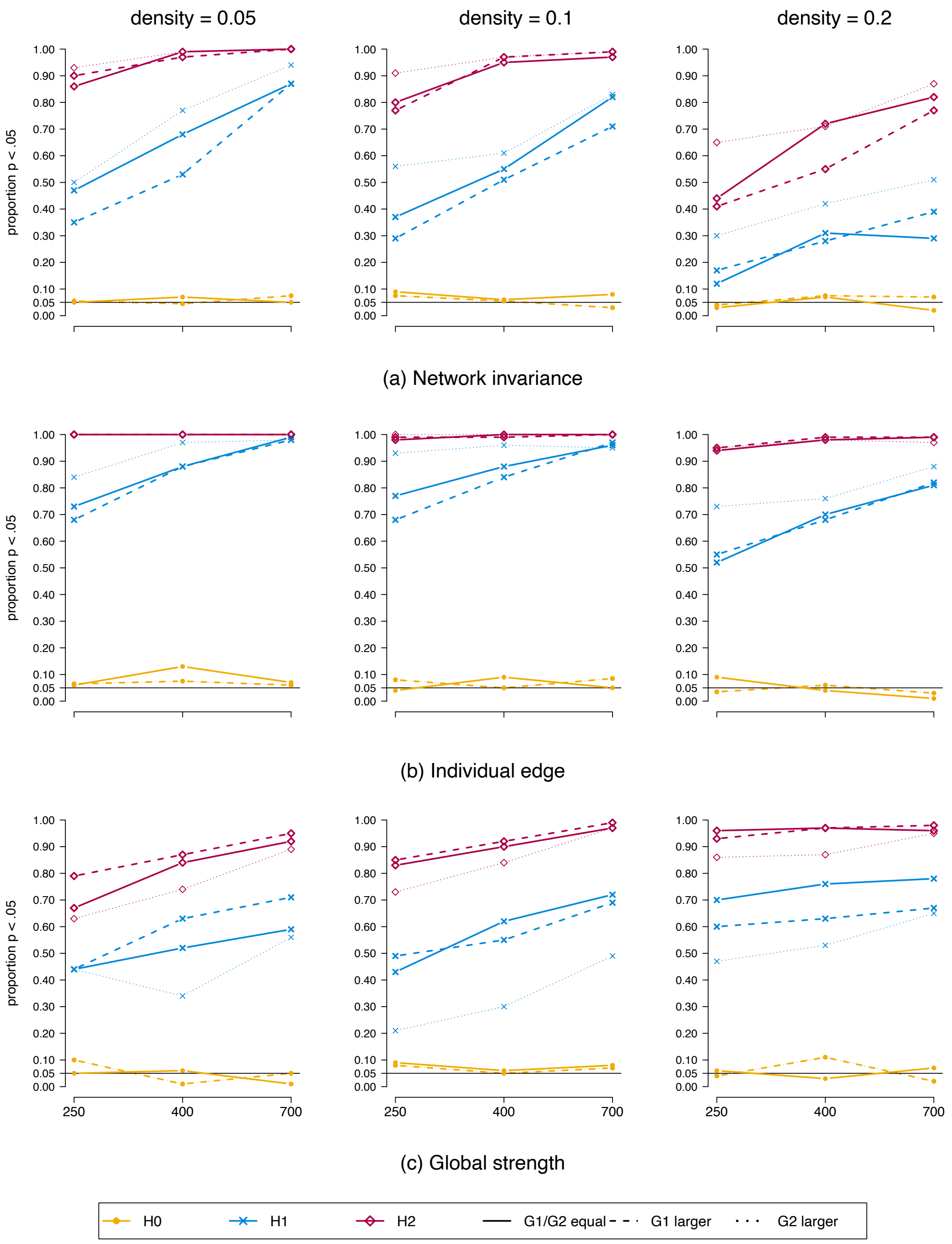
(b) Individual edge

(c) Global strength

Note. Simulation study results for tests on invariance of (a) network structure, (b) an individual edge, and (c) global strength test with continuous data (main analysis). The x-axes display sample size, whereas the y-axes display proportion of *p*-values < 0.05. All three tests were applied to simulated data under the null hypothesis of no difference (H_0, yellow) and under the alternative hypotheses that there is a difference to a certain degree (H_1, blue; and H_2, red). Data were simulated from networks with density 0.5 and different numbers of variables (10: dotted lines, 20: dashed lines, 30: solid lines). The networks were estimated with hyperparameter γ = 0 and data were simulated with equal sample sizes.

Figure S4

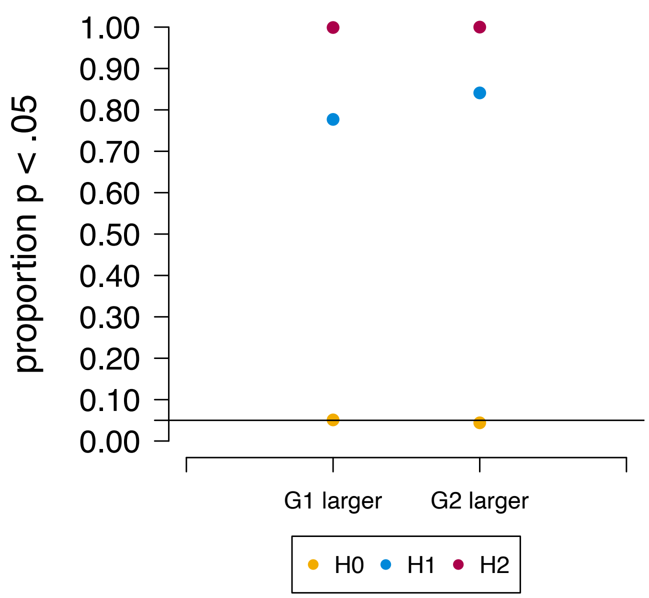
Performance of NCT with binary data with equal and unequal groups (smaller setup)



Note. Simulation study results of the tests on invariance of (a) network structure, (b) an individual edge, and (c) global strength test with binary data (smaller setup). The x-axes display sample size, whereas the y-axes displays proportion of p-values < .05. All three tests were applied to simulated data under the null hypothesis of no difference (H_0, yellow) and under the alternative hypotheses that there is a difference to a certain degree (H_1, blue; and H_2, red). Data were simulated from networks with density 0.05 (left panels), 0.1 (middle panels), and 0.2 (right panels), and 36 variables. The data were simulated with equal (solid line) and unequal sample sizes. In the unequal condition, the largest sample was generated with either network G_1 (dashed line), or with network G_2 (dotted line).

Figure S5

Performance of NCT with settings corresponding to empirical example data



Note. Simulation study results for tests on invariance of an individual edge with continuous data and settings close to the empirical networks: number of nodes = 10, density of G_1 = 0.5, sample sizes = 500/1000, and hyperparameter γ = 0.5. The test was applied to simulated data under the null hypothesis of no difference (H_0, yellow) and under the alternative hypotheses that there is a difference to a certain degree (H_1, blue; and H_2, red). The x-axes displays the two conditions of unequal sample sizes: 1) the data generated with network $G_1$was 2 times larger than that with network G_2 (G_1larger), and 2) the data generated with network $G_2$was 2 times larger than that with network G_1 (G_2 larger). The y-axes display proportion of p-values < 0.05.