

University of Groningen

## Symptom network models in depression research

van Borkulo, Claudia Debora

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2018

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

van Borkulo, C. D. (2018). *Symptom network models in depression research: From methodological exploration to clinical application*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## APPENDIX **E**

### A TUTORIAL ON R PACKAGE NETWORKCOMPARISONTEST

---

Adapted from:

[https://cran.r-project.org/web/packages/NetworkComparisonTest/  
NetworkComparisonTest.pdf](https://cran.r-project.org/web/packages/NetworkComparisonTest/NetworkComparisonTest.pdf)

This tutorial provides an explanation of R package NetworkComparisonTest (Van Borkulo, Epskamp, & Milner, 2016). This Network Comparison Test (NCT) is a permutation based hypothesis test, which is suited for Gaussian and binary data. It assesses the difference between two networks based on several invariance measures (network structure invariance, global strength invariance, edge invariance). Currently, NCT is suited for comparison of independent as well as dependent samples. Concerning dependent samples, only comparing one group which is measured twice is implemented (e.g., a group of patients with a pre- and post-treatment).

Network structures are estimated with  $\ell_1$ -regularized partial correlations (EBICglasso for Gaussian data; Epskamp et al., 2012) or with  $\ell_1$ -regularized logistic regression (eLasso for binary data Van Borkulo et al., 2014).

## E.1 Introduction

Research in which the network approach is used has recently shifted from a descriptive stance to a more comparative stance (Bringmann et al., 2013; Pe et al., 2015; Wigman et al., 2015). The NetworkComparisonTest package provides a statistical tool to do so, by allowing direct comparison of two networks. This procedure combines advanced methodology for inferring network structures from large empirical, cross sectional datasets (Epskamp et al., 2012; Van Borkulo et al., 2014) with permutation testing. Currently, NCT evaluates three hypotheses that are typically relevant in network analysis: (1) *invariant network structure*, (2) *invariant edge strength*, and (3) *invariant global strength*. The first hypothesis, concerns the structure of the network as a whole, and states that this structure is completely identical across subpopulations. Differently stated, the distributions of edge weights are compared, similar to a Kolmogorov-Smirnoff test. The second hypothesis zooms in on the difference in strength of a specific edge of interest. The third hypothesis says that, although networks may differ in structure, the overall level of connectivity is equal across groups. See Chapter 5, for an extensive explanation of the test statistics, which accompany these three hypotheses.

### E.1.1 Real data to illustrate NCT

In this tutorial, the arguments and output of the function is explained and illustrated with data of the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPUD; Kendler & Prescott, 2006; Prescott et al., 2000). For this tutorial, we use the data of presence/absence of the 14 disaggregated symptoms of MDD for at least 5 days during the previous year of 8973 individuals from the population. To illustrate NCT, we divided the data into males ( $n = 5091$ ) and females ( $n = 3884$ ). This Chapter is an extended version of <https://cran.r-project.org/web/packages/NetworkComparisonTest/NetworkComparisonTest.pdf>.

To investigate differences in network structure between males and females, we can first inspect the networks visually. Judging from Figure E.1, there are no clear differences between the two networks. In men, depressed mood is more strongly associated with feeling worthless and it is associated to weight loss, whereas it is not in women.

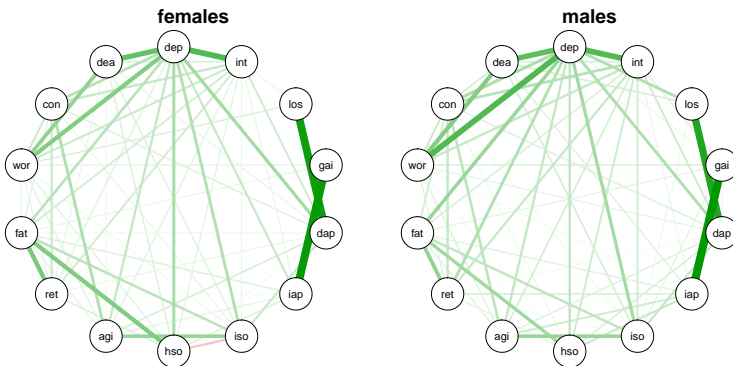


FIGURE E.1. The network structures of females (left panel;  $n = 3884$ ) and males (right panel;  $n = 5091$ ) of the VATSPUD study. Estimation is performed with `IsingFit()` and  $\gamma = .25$ . dep - depressed mood; int - loss of interest; los - weight loss; gai - weight gain; dap - decreased appetite; iap - increased appetite; iso - insomnia; hso - hypersomnia; ret - psychomotor retardation; agi - psychomotor agitation; fat - fatigue; wor - feelings of worthlessness; con - concentration problems; dea - thoughts of death.

## E.2 Arguments

The main function of package `NetworkComparisonTest` is function `NCT()`, which has several arguments:

```
NCT(data1, data2, gamma, it=100, binary.data=FALSE,  
paired=FALSE, weighted=TRUE, AND=TRUE, test.edges=FALSE, edges,  
progressbar=TRUE)
```

In this section, the arguments are explained and the function of some arguments is illustrated with our two groups of individuals (males and females) from the VATSPUD study.

### **data1/data2**

The first two arguments (`data1` and `data2`) are the data of the groups to be compared. Both datasets have to contain cross-sectional data. The dimension of the matrices is  $nobs \times nvars$ , with  $nobs$  (the number of observations) as rows and  $nvars$  (the number of variables) as columns.

### **gamma**

A single value between 0 and 1. When not entered,  $\gamma$  is set to .25 for binary data and to .50 for gaussian data. See Appendix D (D.3) for a more elaborate explanation of effect of adjusting this argument.

### **it**

The number of iterations (permutations) to create a reference distribution (see Section 5.2.3 for an explanation of the reference distribution). The default value is 100.

### **binary.data**

Logical. Can be `TRUE` or `FALSE` to indicate whether the data is binary or not. The default value is `FALSE`, which ensures the data is handled as gaussian. The VATSPUD data that we use to illustrate `NCT` is binary, hence, to compare males and females in this study, the following code should be used.

```
set.seed(1)
res <- NCT(data.females, data.males, binary.data=TRUE)
```

The results of the analysis is assigned to *object* `res`. This object contains the output of NCT. The output is explained in detail in the next section. Note that with this code, a permutation test with 100 iterations (default) is performed, which might be too low considering the large number of observations in our groups ( $n = 5091$  and  $n = 3884$ ).

### **paired**

Logical. Can be TRUE or FALSE to indicate whether the samples are dependent or not — the default value is FALSE. If `paired` is TRUE, relabeling is performed within each pair of observations. If `paired` is FALSE, relabeling is not restricted to pairs of observations. Note that, currently, only comparing one group which is measured twice (e.g., a group of patients with a pre- and post-treatment) is implemented. In the nearby future, NCT will be extended to allow for comparing group-level networks of two groups with ESM data.

### **weighted**

Logical. Can be TRUE or FALSE to indicate whether the networks to be compared should be weighted or not. If not, the estimated networks are dichotomized. Defaults to TRUE.

### **AND**

Logical. Can be TRUE or FALSE to indicate whether the AND- or the OR-rule should be used to define the edges in the network. Defaults to TRUE. Only necessary for binary data. See Section D.2 in the Appendix for a more elaborate explanation of effect of adjusting this argument.

### **test.edges**

Logical. Can be TRUE or FALSE to indicate whether or not differences in individual edges should be tested. Defaults to FALSE. When TRUE, you have to specify which edges you want to test with the next argument `edges`.

## **edges**

Character or list. When “all”, differences between all individual edges are tested. When provided a list with one or more pairs of indices referring to variables, the provided edges are tested. A Holm-Bonferroni correction is applied to control for multiple testing. To test all edges, the following code can be used.

```
set.seed(1)
res2 <- NCT(data.females, data.males, binary.data=TRUE,
test.edges=TRUE, edges="all")
```

 Again, see the next Section E.3, for

how to retrieve the results in the output of NCT.

## **progressbar**

Logical. Should the progressbar be plotted in order to see the progress of the estimation procedure? Especially with binary data, which involves node-wise regressions, it can take a while before all permutations are performed. With the progressbar, you can keep track of the progress of the analysis. Defaults to TRUE.

## **E.3 Output**

NCT returns an *NCT* object that contains several items. Using the code above, the output is stored in an object with the name `res` and `res2`. Since `res2` contains the most elaborate output, we will use that to explain the content of the output.

### **glstrinv.real**

The difference in global strength between the networks of the observed (real) data sets. This is the actual test statistic  $S$  (see Chapter 5) that should be reported. In our example, the difference in global strength  $S$  is 3.72. This can be retrieved with `res2$glstrinv.real`.

**glstrinv.perm**

The difference in global strength between the networks of the permuted data sets. When the number of iterations is 100, `res2$glstrinv.perm` will contain 100 values.

**glstrinv.sep**

The separate global strength values of the individual networks. In our example, `res2$glstrinv.sep` reveals that the females' network has a global strength of 41.97 and the males' network 45.69.

**glstrinv.pval**

The  $p$  value resulting from the permutation test concerning difference in global strength.

`res2$glstrinv.pval` reveals that, although the males have a higher global strength, this difference is not significant ( $S = 3.72$ ,  $p = .26$ ).

**nwinv.real**

The value of the maximum difference  $M$  in any of the edge weights of the observed networks. In our example, `res2$nwinv.real` shows that  $M = .99$ .

**nwinv.perm**

The values of the maximum difference in edge weights of the permuted networks.

**nwinv.pval**

The  $p$  values resulting from the permutation test concerning the maximum difference in edge weights. When `it=100`, there will be 100  $p$  values that form the reference distribution.

**edges.tested**

The pairs of variables between which the edges are called to be tested. Only if `test.edges = TRUE`.



**einv.real**

The value of the difference in edge weight of the observed networks (multiple values if more edges are called to test). Only if `test.edges = TRUE`.

**einv.pvals**

The Holm-Bonferroni corrected  $p$  values per edge from the permutation test concerning differences in edges weights. Only if `test.edges = TRUE` and only for the edges provided in argument `edges.tested`. In our example, `res2$einv.pvals` shows that one edge differs significantly in males and females: the edge between depressed mood (`dep`) and weight loss (`los`). This edge is absent in the males, but present in females ( $p < .01$ ).

**einv.perm**

The values of the difference in edge weight of the permuted networks. Only if `test.edges = TRUE`.

**E.4 Plotting of NCT results**

Results can also be plotted. The permutation test results in a reference distribution of test statistics under the relevant null hypothesis. NCT is accompanied by a plotting function to visualize the results.

In our example of males and females in the VATSPUD study, the results of the network structure and global strength invariance test can be plotted with the following code:

```
plot(res2, what="network")
plot(res2, what="strength")
```

The argument `what` can be used to indicate which statistic to be plotted. Figure E.2 shows the reference distributions — created by permutations of the data — with which the test statistics  $M$  (i.e., the maximum difference in edge strength of the two networks) and  $S$  (i.e., the difference in global strength) can be evaluated.

The distribution(s) of the edge strength invariance test can be plotted with `plot(res2,`

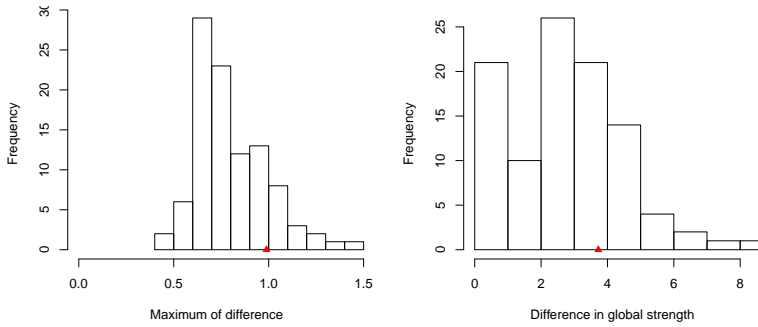


FIGURE E.2. Reference distributions of two of the three test statistics based on the VATSPUD data: the maximum difference in edge strength (left panel) and the difference in global strength (right panel). The red triangle indicates the test statistic based on the observed (real) data.

what="edge"). When more than one edge was tested, more than one plot will be returned (i.e., a plot for each edge tested).

