PSYCHOMETRIKA—VOL. 87, NO. 1, 107–132
MARCH 2022
https://doi.org/10.1007/s11336-021-09765-2

CONNECT: A NOVEL NETWORK APPROACH FOR INVESTIGATING THE CO-OCCURRENCE OF BINARY PSYCHOPATHOLOGICAL SYMPTOMS OVER TIME

NADJA BODNER

KU LEUVEN (UNIVERSITY OF LEUVEN)

LAURA BRINGMANN

UNIVERSITY OF GRONINGEN

FRANCIS TUERLINCKX

KU LEUVEN (UNIVERSITY OF LEUVEN)

PETER DE JONGE

UNIVERSITY OF GRONINGEN

EVA CEULEMANS

LEUVEN (UNIVERSITY OF LEUVEN)

Network analysis is an increasingly popular approach to study mental disorders in all their complexity. Multiple methods have been developed to extract networks from cross-sectional data, with these data being either continuous or binary. However, when it comes to time series data, most efforts have focused on continuous data. We therefore propose ConNEcT, a network approach for binary symptom data across time. ConNEcT allows to visualize and study the prevalence of different symptoms as well as their co-occurrence, measured by means of a contingency measure in one single network picture. ConNEcT can be complemented with a significance test that accounts for the serial dependence in the data. To illustrate the usefulness of ConNEcT, we re-analyze data from a study in which patients diagnosed with major depressive disorder weekly reported the absence or presence of eight depression symptoms. We first extract ConNEcTs for all patients that provided data during at least 104 weeks, revealing strong inter-individual differences in which symptom pairs co-occur significantly. Second, to gain insight into these differences, we apply Hierarchical Classes Analysis on the co-occurrence patterns of all patients, showing that they can be grouped into meaningful clusters. Core depression symptoms (i.e., depressed mood and/or diminished interest), cognitive problems and loss of energy seem to co-occur universally, but preoccupation with death, psychomotor problems or eating problems only co-occur with other symptoms for specific patient subgroups.

Key words: binary data, network analysis, time series, depression, individual differences.

1. Introduction

Network analysis is an increasingly popular approach to study various psychological concepts like attitudes (e.g., Dalege et al., 2016), beliefs (e.g., Brandt et al., 2019), and personality (e.g.,

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11336-021-09765-2.

Correspondence should be made to Nadja Bodner, Quantitative Psychology and Individual Differences Research Group, Faculty of Psychology and Educational Studies, KU Leuven (University of Leuven), Tiensestraat 102, Box 3713, 3000 Leuven, Belgium. Email: nadja.bodner@kuleuven.be

© 2021 The Psychometric Society

Cramer et al., 2012). The approach has especially been advocated and used to investigate mental disorders in all their complexity, based on the hypothesis that such disorders can be conceptualized as systems of connected symptoms (Borsboom, 2008, 2017; Borsboom & Cramer, 2013; Cramer et al., 2010; Fried et al., 2017). The symptoms constitute the nodes of the obtained networks, and the (partial) relations between the symptoms are captured by the edges that connect the nodes.

Network analysis is most often applied to cross-sectional data, where it allows to obtain nomothetic insights in how symptoms relate to each other across individuals (e.g., Boschloo et al., 2015; Cramer et al., 2016; Isvoranu et al., 2016). To this end, multiple cross-sectional methods have been developed for both continuous and binary symptom data (Epskamp, Borsboom, et al., 2018; Fried et al., 2016; van Borkulo et al., 2014).

As network structures have been found to vary hugely across persons (Bosley et al., 2020; Bringmann et al., 2013; de Vos et al., 2017; A. J. Fisher et al., 2017; Wright & Woods, 2020), it is useful to complement cross-sectional analyses with person-specific networks based on time-series data. Such person-specific networks shed light on the within-person relations between symptoms across time that are strongly clinically relevant. However, when it comes to methodological development of techniques to obtain such idiosyncratic, person-specific networks from time series data, efforts have mostly focused on continuous data (e.g., Bringmann et al., 2013; Bulteel, Tuerlinckx, et al., 2018; Epskamp, Waldorp, et al., 2018; Hamaker et al., 2018). However, binary data are also often encountered in psychopathology research, as assessment and treatment in clinical practice are based on the presence or absence of the DSM or ICD symptoms. So far, binary data have mostly been handled through *model-based approaches* (e.g., Haslbeck & Waldorp, 2020; van Borkulo et al., 2014). Although these approaches have many promising and appealing features, they also come with some challenges as we will discuss in the next section. Therefore, we argue that there is room for other approaches to complement and enrich the arsenal of tools for studying binary time-series data, without claiming that one set of tools is better than the other.

This paper has a twofold aim. First, we will propose ConNEcT, a new network approach for analyzing binary symptom data across time. Building on earlier work of Bodner et al. (2018), ConNEcT allows to visualize and study the prevalence of different symptoms as well as their co-occurrence across time in a person-specific network figure. The strength of the co-occurrence of symptoms is measured by means of a contingency measure and can be tested with a permutation-based significance test that was validated in extensive simulations (Bodner et al. 2020). Importantly, this test accounts for the serial dependence within each symptom (i.e., the likelihood of a symptom depends in many cases on its presence or absence at the previous measurement occasion). As shown by Bodner and colleagues (in press), disregarding such serial dependence in the testing procedure may lead to type one errors. ConNEcT differs from the above mentioned model-based approaches in that it is a descriptive approach that investigates pairwise co-occurrence, rather than conditional relations, and does not impose model-based assumptions about how these connections come about. Second, we aim to illustrate the usefulness of ConNEcT by re-analyzing the data collected by Hosenfeld et al. (2015). These authors asked 267 patients diagnosed with major depressive disorder to weekly report the absence or presence of eight depression symptoms. ConNEcT will yield a separate symptom co-occurrence network for each of these patients. We expect that these networks display strongly heterogeneous symptom co-occurrences across patients. Indeed, earlier work has already shown that depression plays out very differently in individuals. For instance, Fried and Nesse (2015) found no less than 1000 unique symptom patterns, indicating strong individual differences in prevalence and co-occurrence of symptoms. Although these observed differences do not necessarily imply a difference in the underlying data generating model (Marsman et al. 2019), studying these different co-occurrence patterns will shed further light on how different types of symptom connections characterize the vicious circle we call depression (Borsboom, 2017; Borsboom & Cramer, 2013; Fried et al., 2017). To gain insight into these between-person differences in symptom co-occurrence, we will apply hierarchical classes analy-

sis (HICLAS; De Boeck & Rosenberg, 1988) on the person-specific networks, showing that they can be grouped into meaningful clusters.

The remainder of this paper is organized as follows: First, the ConNEcT approach is introduced. We will explain the method and show how to construct a person-specific co-occurrence network for one example patient. Herewith we will also introduce the significance test and show how applying it prunes the network of the example patient. Next, we will discuss how HICLAS can be used to group the heterogeneous idiographic networks into a few clusters. We end this theoretical section by shedding light on some differences between ConNEcT and logistic regression approaches. In the illustrative application section, we will apply the ConNEcT approach as well as HICLAS to the data from the Hosenfeld et al. (2015) study. The final section contains a discussion of our methods and some directions for future research.

2. The ConNEcT Approach

2.1. Building a ConNEcT for an Individual Person

In this section, we describe how to build a ConNEcT for the symptoms of a single person, in four steps. First, we compute the prevalence and serial dependence of each of the symptoms. Second, we quantify the co-occurrence of each symptom pair. Third, we apply a significance test on the co-occurrence values to determine which of these values are larger than expected by chance given the prevalence and serial dependence of the two symptoms. Fourth, we draw the associated network, showing either all co-occurrence values (raw network), or only the significant ones (pruned network). We will now explain each of these steps in more detail, using the data of one example patient from the Hosenfeld study. Figure 1 shows the weekly presence/absence of the eight depression symptoms across 145 weeks, for this patient (Fig. 1a), the raw network with all links (Fig. 1b) and the pruned network with the significant links only (Fig. 1c).

Step 1 Quantifying prevalence and serial dependence From Fig. 1a, it can, for instance, be derived that the core symptoms (i.e., depressed mood and/or diminished interest) were present (i.e., elevated line) in the first 68 weeks, in week 77 until week 103 and finally in the last eight weeks, while they were absent between weeks 69 and 76 and between weeks 104 and 138. We will indicate that symptom X is present in week t as $X_t = 1$, whereas its absence in this week is denoted as $X_t = 0$. Moreover, we also see that all symptoms show strong serial dependence, in that symptoms are likely to last longer than one week.

The *prevalence* of each symptom X is quantified in terms of its relative frequency $p_1$over weeks, where $p_1 = p(X_t = 1)$. The obtained values in Table 1 show that eating problems were never reported and that core symptoms, cognitive problems, sleeping problems and psychomotor problems were most prevalent. The *serial dependence* of the symptoms can be quantified in terms of two conditional probabilities. Specifically, we compute the probability of observing a '1' in a week, conditional upon the value in the previous week, yielding $p_{1|1}$ (short for $p(X_t = 1|X_{t-1} = 1)$) and $p_{1|0}$ (short for $p(X_t = 1|X_{t-1} = 0)$). If both conditional probabilities are equal and hence do not differ from the relative frequency p_1, there is no indication of serial dependence. The more both conditional probabilities deviate, the stronger the serial dependence. From Table 1, we conclude that in our example data all symptoms show serial dependence.

Step 2 Quantification of co-occurrence The ConNEcT approach quantifies the strength of the co-occurrence of each pair of symptoms across time by means of a contingency measure. Frequently used contingency measures for binary data are, for example, the proportion of agreement and its complement the Hamming distance (Hamming, 1950), the (log) odds ratio, Cohen's kappa, Wampold's Transformed kappa (Bakeman et al., 1996; Holloway et al., 1990), the phi-coefficient and Yule's Q (Bakeman et al., 1996; Bakeman & Quera, 2011), but also the Forbes coefficient and the tetrachoric correlation have been proposed (Salvatore et al., 2019). In line with some

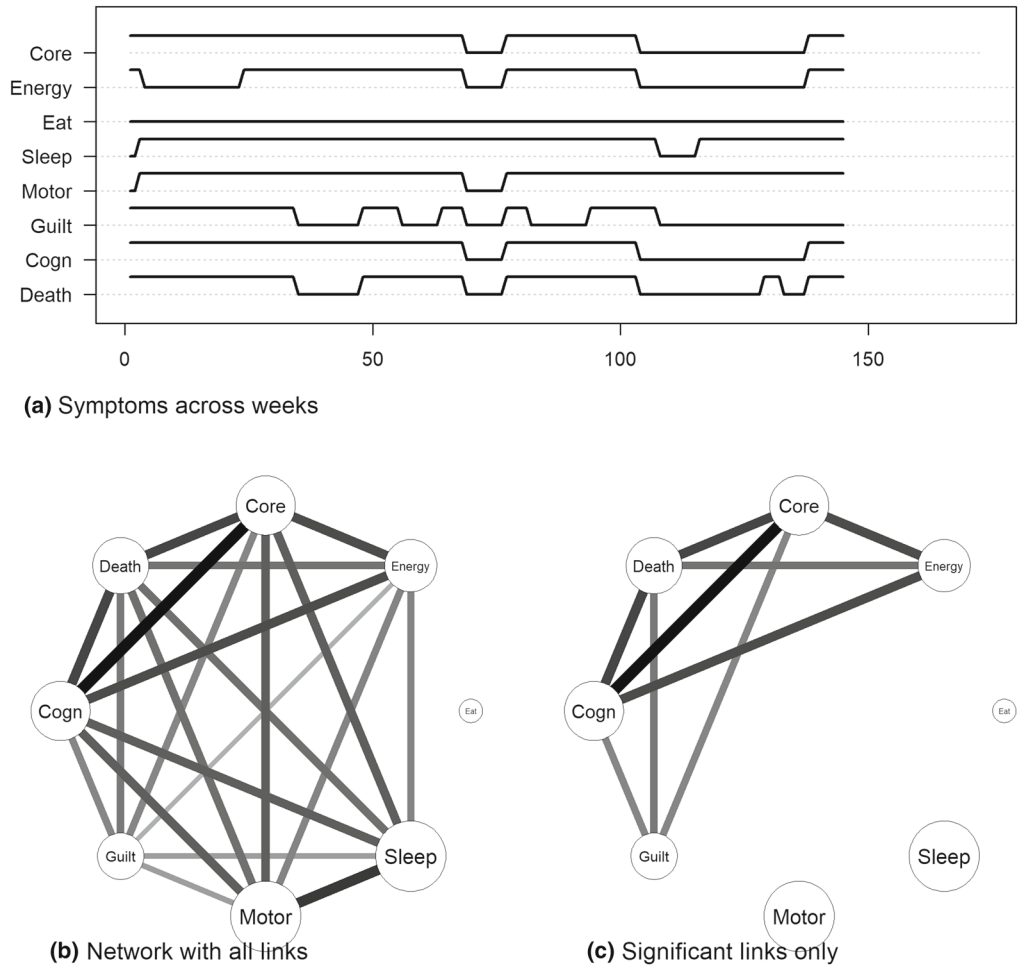


FIGURE 1.
Reported weekly presence of eight depression symptoms for the example patient and the two associated ConNEcTs. **a** Patient's weekly reports of the eight symptoms (**Core** symptoms, lack of **energy**, **eat**ing problems, **sleep**ing problems, psycho**motor** problems, feelings of **guilt**, **cogn**itive problems and preoccupation with **death**), where the base lines indicate absence and the elevated lines presence of a symptom; **b** raw symptom co-occurrence network of the example patient, containing all links; **c** pruned symptom co-occurrence network of the example patient showing significant links only. In panels b and c, the node size is adapted to the relative frequency of the symptoms and the width and saturation of the edges to the size of the corresponding Jaccard value.

recent papers (Bodner et al., 2018; Brusco et al., 2019; Van keer et al., 2019), we used the Jaccard similarity index, because it has some interesting characteristics.

The Jaccard index can be defined as follows (Jaccard, 1901, 1912):

$$Jac = \frac{n_{11}}{n_{11} + n_{10} + n_{01}},$$

where n_{11} denotes the number of weeks that both symptoms are reported as present, and n_{10} and n_{01} the number of weeks that only the first or only the second symptom is reported as present. The Jaccard index thus boils down to the ratio of the weeks in which both symptoms are shown over those where at least one of them is shown. The Jaccard index is symmetric, implying that it

TABLE 1.
Relative frequencies, serial dependencies and the Jaccard values of the example patient.

	rel.freq	Serial dependencies		Jaccard values						
	p_1	$p_{1\|1}$	$p_{1\|0}$	Energy	Eat	Sleep	Motor	Guilt	Cogn	Death
Core	0.71	0.98	0.05	**0.81**	0	0.74	0.74	**0.58**	**1.00**	**0.84**
Energy	0.57	0.96	0.05		0	0.59	0.59	0.39	**0.81**	**0.65**
Eat	0.00	NaN	0.00			0.00	0.00	0.00	0.00	0.00
Sleep	0.93	0.99	0.20				0.89	0.47	0.74	0.67
Motor	0.93	0.99	0.20					0.47	0.74	0.67
Guilt	0.46	0.92	0.05						**0.58**	**0.63**
Cogn	0.71	0.98	0.05							**0.84**
Death	0.65	0.96	0.08							

$p_{1|1}$: short for $p(X_t = 1|X_{t-1} = 1)$; conditional probability that a symptom is reported given that it had been reported the week before, $p_{1|0}$ short for $p(X_t = 1|X_{t-1} = 0)$; conditional probability that a symptom is reported given that it had not been reported the week before. Significant Jaccard values are printed in bold.

does not play any role which of the two symptoms is considered as first and which as second. The obtained Jaccard values range from 0 (no co-occurrence across time) to 1 (perfect co-occurrence).

Like many other measures, the value of the Jaccard index depends on the relative frequency of the symptoms, with higher relative frequencies leading to higher Jaccard values (Bodner et al. 2020; Salvatore et al., 2019). The measure differs, however, from a number of contingency measures in that no expected value is subtracted, which we consider an asset since the computation of the expected value varies from one measure to the other anyway (e.g., Cohen's kappa versus Scott's pi; Cohen, 1960; Scott, 1955). Moreover, the Jaccard index differs from other popular similarity indices such as the proportion of agreement, the Hamming distance or Cohen's Kappa, in that weeks in which both symptoms are absent (n_{00}) are not taken into account. Treating the co-occurrence and co-absence of symptoms as exchangeable might lead to the erroneous conclusion that the co-absence of sleeping problems and psycho-motor problems might be indicative for a certain type of depression. In case of co-occurrence of both symptoms, this conclusion obviously makes sense. Although the co-absence of symptoms might be meaningful in other contexts, we do not think that is the case here, and therefore only account for co-occurrence.

Table 1 shows the obtained Jaccard values for all symptom pairs of our example patient. The values for pairs that involve eating problems amount to zero, since this symptom is never present. We observe high values for the links between cognitive and core problems (1, indicating perfect co-occurrence) and between psychomotor and sleeping problems (.89). The link between feelings of guilt and loss of energy is the weakest (.39).

Step 3 Applying a significance test The interpretation of Jaccard values, as well as the values of many other similarity indices (see Bakeman & Quera, 2011; Bodner et al. 2020), is sometimes difficult because its values also reflect the relative frequency of the symptoms under consideration. Therefore, Jaccard values for different symptom pairs can only be compared, if the underlying relative frequencies are equal. For instance, we see in Table 1 that the relative frequency of guilt (0.46) is much lower than that of psychomotor problems (0.93). And indeed all Jaccard values that involve psychomotor problems are higher (e.g., with cognitive problems: 0.74) than those between guilt and the same symptom (e.g., cognitive problems: 0.58). It is therefore not a good idea to compare the two Jaccard values directly (e.g., retain the biggest ones, Brusco et al., 2019).

Moreover, although less documented and discussed, the values of similarity indices are also impacted by serial dependence. For instance, if two symptoms co-occur in a specific week and both show strong serial dependence, chances are high that both symptoms will also be present

in the next week, and vice versa. This implies that strong serial dependence leads to much wider sampling distributions (Bodner et al. 2020). Therefore, we propose to run a one-sided significance test that builds a sampling distribution of the co-occurrence measure under the null hypothesis that the two symptoms do not systematically co-occur. This significance test that was validated in extensive simulations by Bodner et al. (2020) is a permutation test, in that it generates the sampling distribution by permuting the original data (for a general introduction to the statistical foundation of permutation tests, see e.g., Edgington & Onghena, 2007; R. A. Fisher, 1949; Good, 2000). The challenge here was to determine how the original data should be permuted, to remove systematic co-occurrence (given the null hypothesis), but keep the observed relative frequency and serial dependence. This can be achieved by implementing the segment shuffling approach from Moulder et al. (2018), in which the time series of each symptom is cut into ten roughly equal-sized segments (for a similar idea in a cross-validation approach, see Bulteel, Mestdagh, et al., 2018) of 10–15 adjacent weeks. The order of these ten segments is then randomly permuted for each symptom separately (see Fig. 2). This way, the pairwise association between both symptoms is broken down, shedding light on which co-occurrence values can be observed without it. Importantly, using segment shuffling, the relative frequencies are held constant (all 0s and 1s will be retained, but in a different order) and the serial dependence (and other secondary characteristics of the data) from the original data is largely retained (Bodner et al. 2020; Moulder et al., 2018), because sequences of reported symptoms are largely kept and moved together. This is important, because serial dependence can lead to some accidental and thus ignorable co-occurrence. When serial dependence is disregarded, for instance, by randomly shuffling all weeks (i.e., segments of length 1), this phenomenon leads to an increase in type 1 errors (see results of the simulation study by Bodner et al. 2020), because the pairwise association due to serial dependence is also broken down. A segment shuffling approach, on the other hand, keeps most of the serial dependence intact; hence, the testing result is not affected by the associated accidental co-occurrence.

For each symptom, the random permutation of the segments is repeated 1000 times. Afterward, for each symptom pair the Jaccard value is calculated for all possible pairs of associated permuted symptom data (i.e., $1\ 000^2$ pairs). Finally, the Jaccard value for the original data is compared to the distribution of the permuted Jaccard values that constitutes a sampling distribution under the null hypothesis of no inter-dependence. If the original value is larger than the 95^{th} percentile of the permuted values, the original value is considered to be significantly higher than what we would expect based on chance. Bodner et al. (2020) have shown in an extensive simulation study that the segment shuffling test results in a type 1 error rate that is close to the adopted significance level.

If we scrutinize the obtained p-values for the two links we considered above, motor-cognitive and guilt-cognitive, we see that the Jaccard value of the former is higher (0.74; $p = .0652$) than that of the latter (0.58; $p = .0035$), but only the latter link is significant. The symptoms guilt and cognition thus co-occur more often than could be expected by chance, given their relative frequency and their serial dependence. Figure 3 illustrates this further by presenting the sampling distributions under the null hypothesis for both links, as well as the corresponding significance thresholds and observed Jaccard values. We observe that based on the prevalence and serial dependence of different symptoms, we generally expect the Jaccard values for the Motor-Cognition link to be higher than those for the Guilt-Cognition link, explaining the significance results. In total, only nine symptom pairs co-occur more than expected by chance. These nine pairs include the symptoms loss of energy, cognitive and core problems, feelings of guilt and preoccupation with death.

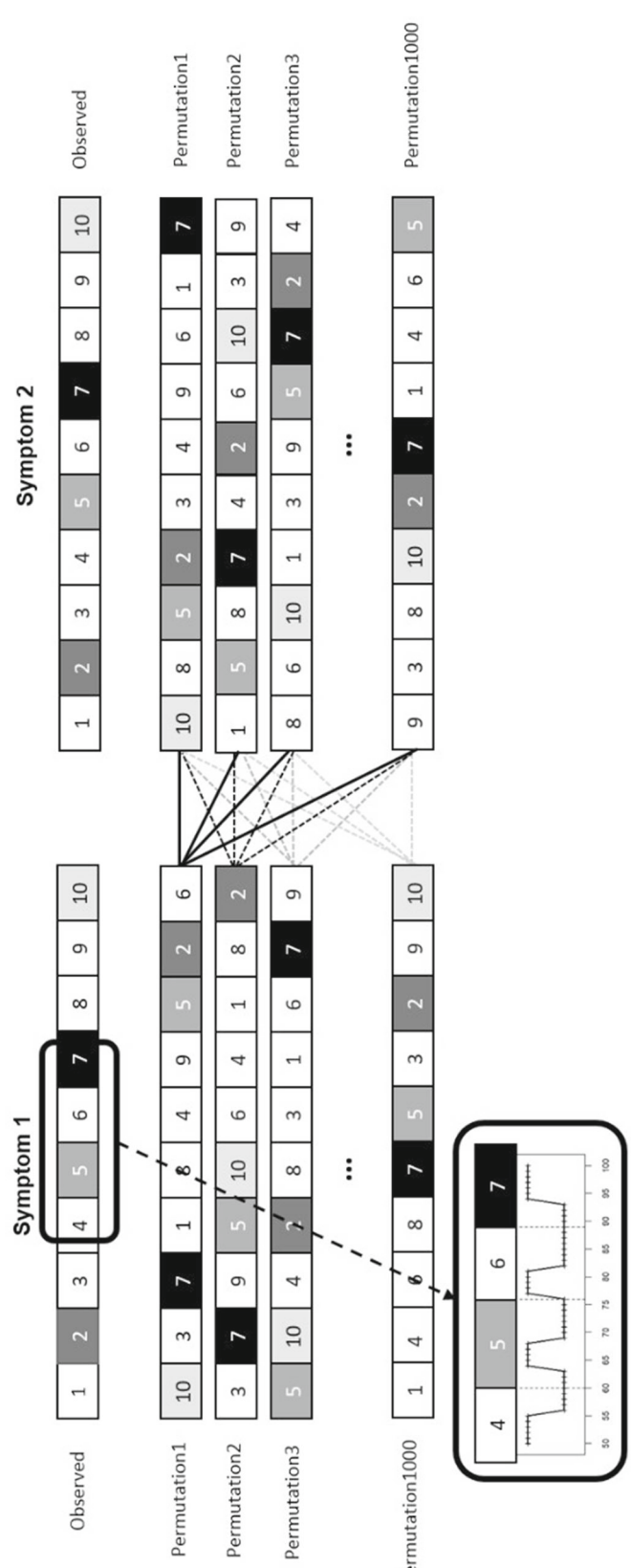


FIGURE 2.

Schematic representation of the segment shuffling approach. The two observed time series (Symptom 1 and Symptom 2) are first cut in ten rather equally sized segments, that each consist of a series of 10–15 adjacent weeks (see small frame, that illustrates the segmenting for the Guilt data of the example patient). These segments are randomly rearranged per symptom, yielding 1000 permuted versions. To distinguish the solutions, we colored the second, fifth, seventh and 10th segment in different shades of gray. For each possible pair of permuted versions, the Jaccard value is calculated.

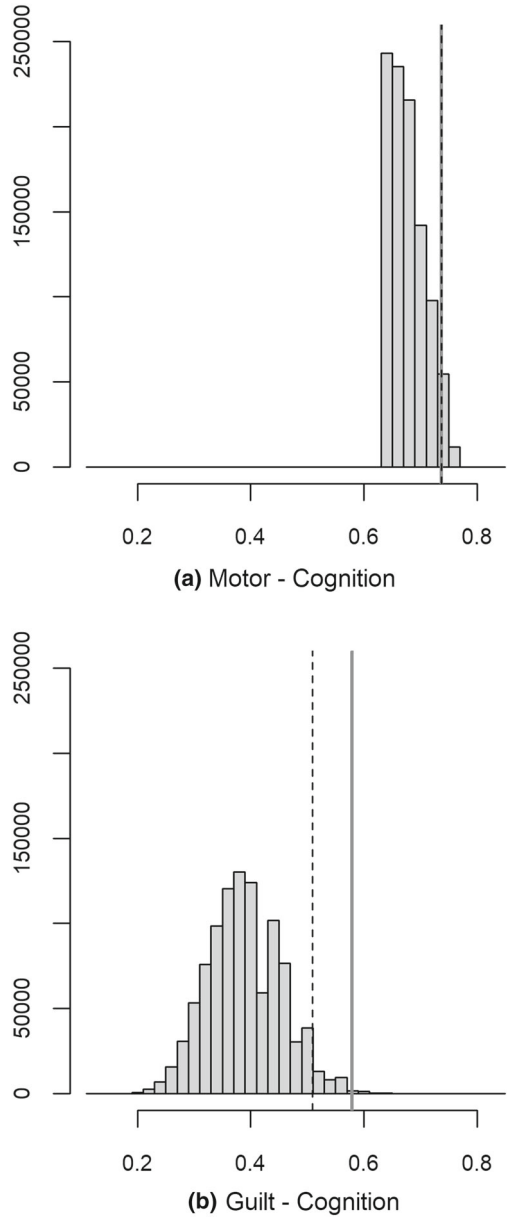


FIGURE 3.
Sampling distributions under the null hypothesis and observed Jaccard values (solid lines) of the links between **a** psychomotor and cognitive problems and **b** feelings of guilt and cognitive problems. The dashed lines indicate the 95% significance threshold of the one-side test. In Fig. 3b the observed value is larger than the significance threshold, indicating a significant link. In Fig. 3a, the observed value is smaller than the threshold, implying that the two symptoms do not co-occur more often than expected by chance.

Step 4 Visualizing the results in a network figure Different symptoms constitute the nodes of the network (see Fig. 1b and c). Their prevalence is depicted in the node size.[1] To indicate co-occurrence, we draw edges between the nodes, the width and saturation of which are proportional to the obtained contingency value. Indeed, Fig. 1b shows the raw co-occurrence network for our example patient, based on the Jaccard values in Table 1. The pruned network (Fig. 1c) only contains the edges for which the associated Jaccard values passed the significance test. Comparing both networks reveals that only the nine links mentioned above (bold in Table 1) were retained. Both networks were drawn using the qgraph package (Epskamp et al., 2012).

2.2. *Clustering ConNEcTs of Multiple Persons Through HICLAS*

If multiple persons reported on their symptoms, ConNEcT will yield a separate significance-pruned network for each person. To summarize the main similarities and differences of these networks, we propose to apply HICLAS (De Boeck & Rosenberg, 1988) on the significance results of all persons. We collect all these results in a binary I persons by J links matrix, indicating all significant links in the person-specific networks with '1' and all non-significant ones with '0.' Denoting the number of symptoms as S, J equals $S(S\text{-}1)/2$.

HICLAS can be conceived as a principal components approach for binary data in which K components are extracted and both the obtained component scores and the component loadings on these K components are binary. The component scores are represented in an I persons by K components matrix **F** (with entries f_{ik}) and the component loadings in an J links by K components matrix **B** (with entries b_{jk}). Like component analysis, the goal is to find those components that allow to reconstruct the original significance-pruned networks as well as possible, that is, the sum of the squared deviations between the observed significance results d_{ij} and the reconstructed results m_{ij} ($i = 1..I$, $j = 1..J$) is minimized. Importantly, HICLAS differs from standard component analysis in that the scores in **F** and the loadings in **B** are combined in a logic rather than algebraic way, when computing the reconstructed significance results m_{ij}. According to the HICLAS model, a person will have a significant link in his/her network, if this link has a loading of '1' on at least one of the components to which the person belongs (i.e., the person has a component score equal to '1'). Formally, this rule can be written as follows:

$$m_{ij} = 1 \Leftrightarrow \exists k : f_{ik} = 1 \wedge b_{jk} = 1$$

where f_{ik} denotes the component score, b_{jk} the component loading, and $\wedge$ the logical *and*. In other words, the reconstructed ConNEcT of a specific person does not contain a particular link if either the person does not belong to any component, or only belongs to component for which the link has a zero-loading on the component.[2]

HICLAS models do not have closed form solutions. Therefore, **F** and **B** are estimated through alternating least squares approaches, where **F** is updated based on a current estimate of **B** through Boolean regression, and vice versa, until the sum of the squared deviations between d_{ij} and m_{ij} does not decrease anymore. Multiple starts (i.e., initializations of **F** and **B**) are used to avoid local minima (Ceulemans et al., 2007).

The fit of a HICLAS model is usually quantified by the sum of the squared deviations between d_{ij} and m_{ij} (i.e., a badness-of-fit criterion BOF). To determine the optimal number of components K, that describes the data well without yielding an overly complex solution, a number of model selection criteria have been discussed, of which applying a variant of the scree test on the BOF

[1] For reasons of readability a minimum node size was applied, so that symptoms that were not shown still were represented in the network.

[2] This HICLAS reconstruction rule is a disjunctive rule; note that also conjunctive versions have been proposed (see Van Mechelen et al., 1995).

value is most popular (Ceulemans & Van Mechelen, 2005; Ceulemans & Kiers, 2006; Wilderjans et al., 2013). The key idea is to look for an elbow in the number of components by *BOF* value plot, as this elbow denotes the number of components after which the decrease in *BOF* levels off. This elbow is quantified in terms of an *st*-criterion (see Wilderjans et al., 2013):

$$st_k = \frac{BOF_{k-1} - BOF_k}{BOF_k - BOF_{k+1}},$$

where k indicates a considered number of components. The k-value that yields the highest st_k-value is retained. Yet, as is often the case in data analysis, elbows might be quite weak or multiple elbows might be present. Hence, interpretability of the obtained **F** and **B**, will often be taken into account in the model selection decision.

As is the case in component analysis, the K retained HICLAS components can be labeled by inspecting the links that uniquely load on a specific components. Moreover, because of the binary nature of the component scores, HICLAS partitions the persons into 2^K mutually exclusive clusters, where K denotes the number of components. To interpret these 2^K clusters, their reconstructed networks can be drawn and compared. Indeed, based on the HICLAS rule, HICLAS yields a reconstructed significance network for each person, with these networks obviously being the same for persons with identical component scores.

2.3. A Comparison of ConNEcT and Model-based Approaches for Binary Data

In this subsection, we will compare ConNEcT to network approaches that are based on specific assumptions about the models that are assumed to underlie the data. The most popular examples thereof are mixed graphical (Haslbeck & Waldorp, 2020) and Ising models (van Borkulo et al., 2014). The parameters of these models can be fit through the IsingFit and mgm packages, making use of logistic regression. This points towards a first difference between ConNEcT and the model-based approaches: ConNEcT is a descriptive approach that stays close to the data. It considers the connections pairwise and quantifies and tests co-occurrence, without making model-based assumptions about how these connections come about. ConNEcT thus ignores that it is very well possible that differences in observed symptom patterns may be consistent with a single data generating model (Marsman et al. 2019). This absence of an underlying model specification further becomes apparent if one realizes that the significance test is a non-parametric permutation test based on reshuffling the original data (see *Step 3: Applying a significance test)*. Not surprisingly, we show in appendix that ConNEcT succeeds quite well in capturing the pairwise connections that result from rather different model specifications (see Appendix).

A second difference between ConNEcT on the one hand and mgm and IsingFit on the other hand pertains to serial dependence in the data and how this is accounted for in the significance test. Indeed, the logistic regression approaches that underlie mgm and IsingFit were developed for cross-sectional data and do not yet account for serial dependence.[3] Therefore, the obtained type 1 error rates are likely to be inflated in case of strong serial dependency, as we show in Appendix. In contrast, as shown in Bodner et al. (2020), ConNEcT controls for serial dependence through the segment shuffling operation in the permutation test.

Third, whereas this paper squarely focuses on the (marginal) pairwise relationships between the symptoms, mgm and Isingfit are applied to study conditional relations, controlling for the presence/absence of all other symptoms. Yet, if we use the underlying logistic regression ideas

[3] It should be noted that mgm also implements regression-based time-series models for binary data as a special case of mixed VAR models, allowing to control for serial dependence by including lagged outcome variables. This approach yields more complicated regression equations including at least two predictor variables, however. Therefore, we did not consider it further.

to investigate simple marginal pairwise relationships, in which the presence/absence of a first symptom is predicted by the presence/absence of a single second symptom, without controlling for other symptoms, we encounter some identification issues. To explain these issues, we first note that in well-conditioned data, such simple logistic regressions will yield the same regression coefficient regardless which of the two symptoms is used as criterion and which as predictor. The obtained regression coefficient then equals the log-odds ratio of both symptoms:

$$LOR = \log\left(\frac{n_{11}n_{00}}{n_{10}n_{01}}\right).$$

For our example patient, running such simple logistic regressions for the symptoms feeling of guilt and cognitive problems yields a regression weight of 2.66, which equals the log-odds ratio, since the odds ratio amounts to 14.37. However, for some other combinations of symptoms, we encountered identification problems, which are at least partly due to so-called (quasi-)complete separation features (Albert & Anderson, 1984). Specifically, as can be derived from the formula of the log-odds ratio, one runs into trouble when one of the symptoms never occurs or never occurs without the other, implying that the denominator of the ratio equals zero. If the numerator is zero (e.g., if the two symptoms never co-occur), the log-odds ratio cannot be calculated either. For our example patient, these identification problems occurred for the link between core symptoms and cognitive problems. Since these two symptoms are always shown together (implying a Jaccard value of 1), the denominator of the odds ratio equals zero. Also, for the link between Core and Energy (where a Jaccard value of .81 indicates high co-occurrence), we ran into identification problems for logistic regression.[4] based approaches: Loss of energy was never shown without core symptoms, implying that the denominator of the odds ratio equals zero. These specific problems only occur for the log odds ratio, while other contingency measures calculate a reasonable value in these standard situations. One can, of course, also think of other situations, for example, cases where symptoms are almost never shown, where most measures will be unreliable.

3. Illustrative Application: A Re-analysis of the Hosenfeld et al. Data

With this section, we aim to illustrate that the ConNEcT approach is able to retrieve meaningful symptom occurrence patterns within persons. Moreover, we will show how the person-specific ConNEcT results can be insightfully summarized by means of HICLAS. To this end, we re-analyzed the weekly assessed symptom data from a study by Hosenfeld et al. (2015).

3.1. Data

The Hosenfeld et al. data were collected from 267 patients, all diagnosed with major depressive disorder (MDD). For details on the sample, see the original paper.[5] These patients were contacted by phone every three months and asked to retrospectively report the weekly presence or absence of eight symptoms that were based on the DSM-IV criteria for depression: depressed mood & diminished interest (Core), eating problems (Eat), sleeping problems (Sleep), psychomotor problems (Motor), loss of energy (Energy), feelings of guilt (Guilt), cognitive problems (Cogn), and preoccupation with death/suicidal ideation (Death).

[4] To circumvent this identification problem solutions have been proposed, which make use of regularized or Bayesian logistic regression (Firth, 1993; Gelman et al., 2008; Heinze & Schemper, 2002; Mansournia et al., 2018).

[5] The Medical Ethics Committee of the University Medical Center Groningen (UMCG) approved the study (No MEC96/02/028c). All participants provided informed consent.

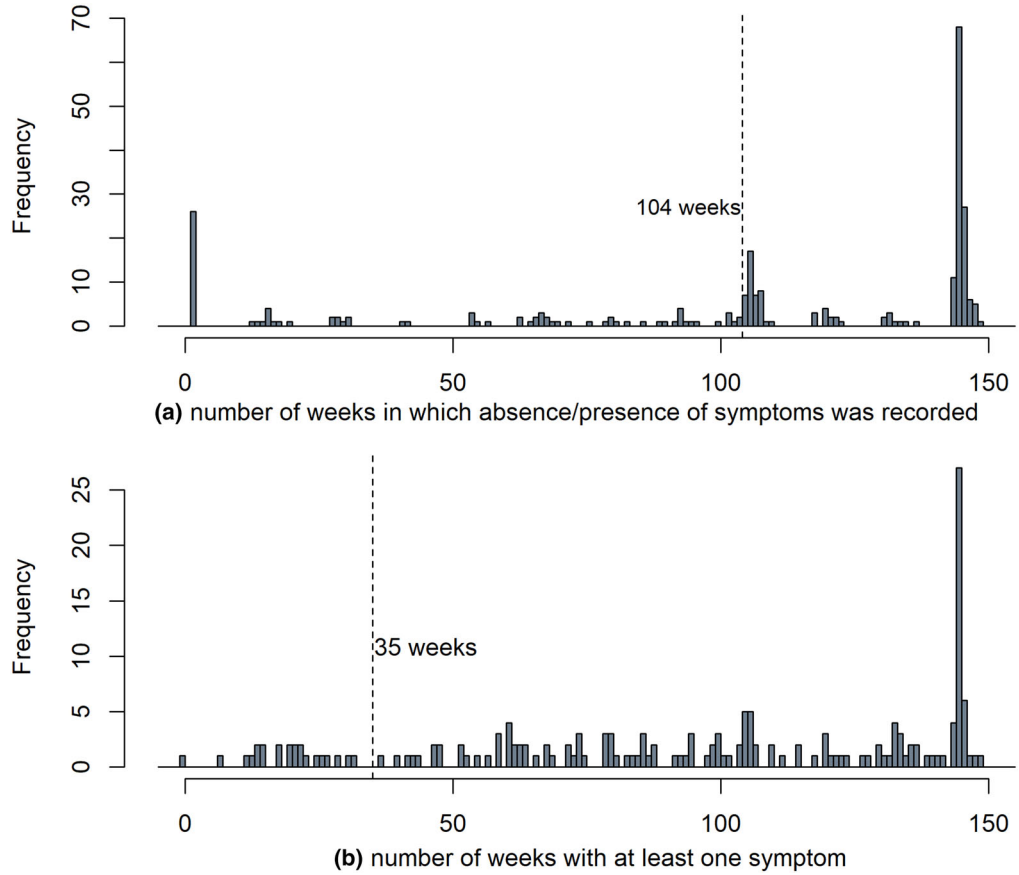


FIGURE 4.
Visualization of the adopted selection criteria. **a** Number of weeks in which the absence/ presence of symptoms was reported; **b** Number of weeks in which at least one symptom was reported.

Out of the 267 patients, we selected all 158 patients that provided data for at least 104 consecutive weeks and showed at least one symptom during at least 35 weeks. As Bodner et al. (2020) showed that the significance test can be performed in a reliable way if 100 measurements are available, we opted for 104 weeks which is equivalent to two years. As shown in panel a of Fig. 4, most of the discarded patients had far less observations. The second criterion, regarding the number of weeks in which at least one symptom is shown, was introduced because it makes no sense to study co-occurrence of symptoms if there are almost never any symptoms present. We opted for 35 weeks, because 35 roughly equals $1/3^{rd}$ of 104 and because we see a gap around that week in the histogram in panel b of Fig. 4.

3.2. Patient-specific ConNEcTs: Number and Nature of Significant Links

Using a significance level of 5%, the total number of significant co-occurrence links per patient ranged between 0 and 25, with a median of 5, indicating a positively skewed distribution (see panel a of Fig. 5). Also, the frequency with which a certain link was significant differed considerably across the links, ranging between present in none of the patients (e.g., some links involving preoccupation with death) and present in 78 individuals (i.e., the link between core symptoms and loss of energy). The number of significant co-occurrence links varied considerably

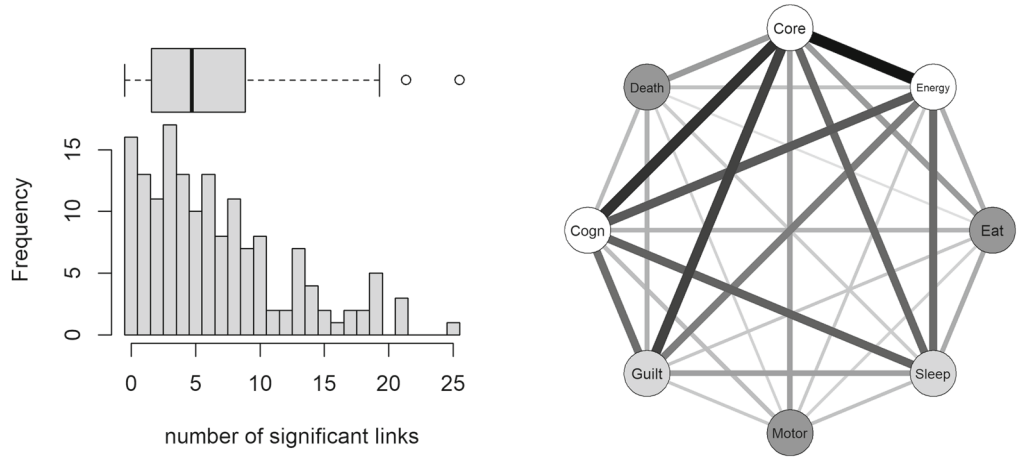


FIGURE 5.
Graphical representation of the number and nature of the significant co-occurrence links. **a** Histogram and boxplot of the number of significant co-occurrence links per patient; **b** network of the significant co-occurrence links, in which the width and saturation of the edges reflect the number of patients for which this link was significant. The colors of the nodes indicate the median number of links per node (white: median = 2; light gray: median = 1; dark gray: median = 0).

TABLE 2.
Number of patients with a specific number of significant co-occurrence links per symptom.

	Number of significant co-occurrence links									
Symptom	0	1	2	3	4	5	6	7	Mean	Median
Core	32	23	33	20	25	15	10	0	2.43	2
Cogn	58	16	22	21	16	17	7	1	2.03	2
Energy	50	23	27	25	14	7	12	0	1.99	2
Sleep	62	20	25	17	20	8	5	1	1.76	1
Guilt	64	22	26	11	13	15	6	1	1.75	1
Eat	80	32	17	13	7	4	5	0	1.16	0
Death	92	22	18	7	9	7	2	1	1.07	0
Motor	98	17	17	9	7	5	3	2	1.03	0

across patients (Fig. 5a), as well as across symptoms (see Fig. 5b and Table 2). The nodes for core symptoms, loss of energy and cognitive problems showed two or more significant links in most of the person-specific networks, whereas eating problems, psychomotor problems and preoccupation with death had no significant co-occurrence links.

3.3. A HICLAS Summary of the Patient-specific ConNEcTs

To group the patients according to their specific significance networks, we conducted HICLAS on the 158 patients by 28 links data (i.e., 8*7/2), extracting one to five components. Since the scree plot did not show a very clear elbow, we compared the interpretability and complexity of the obtained solutions and retained three components.

Component Loadings of the Links

Table 3 shows the loadings of the 28 co-occurrence links on the three HICLAS components. The first components is characterized by five links involving psychomotor symptoms (Motor) and the link between Sleep and Guilt. The second component consists of four links containing

TABLE 3.
Component loadings of the symptom links in the HICLAS model with three components.

Link	Component loadings		
	F1	F2	F3
Core – Motor	1	0	0
Energy – Motor	1	0	0
Sleep – Motor	1	0	0
Sleep - Guilt	1	0	0
Motor – Guilt	1	0	0
Motor - Cogn	1	0	0
Core – Eat	0	1	0
Energy – Eat	0	1	0
Eat – Sleep	0	1	0
Eat – Cogn	0	1	0
Core – Death	0	0	1
Guilt – Death	0	0	1
Cogn – Death	0	0	1
Core – Sleep	1	1	0
Energy – Sleep	1	1	0
Sleep – Cogn	1	1	0
Core – Guilt	1	0	1
Energy – Guilt	1	0	1
Guilt – Cogn	1	0	1
Core - Energy	1	1	1
Core – Cogn	1	1	1
Energy – Cogn	1	1	1
Eat – Motor	0	0	0
Eat – Guilt	0	0	0
Energy – Death	0	0	0
Eat – Death	0	0	0
Sleep – Death	0	0	0
Motor – Death	0	0	0

eating problems (Eat). Three links containing preoccupation with death (Death) load uniquely on the third component. Therefore, we will label the three components as psychomotor problems, eating problems, and pre-occupation with death. Some of the other links have cross-loadings, indicating that they are significant more often. Specifically, three links involving Sleep belong to both the Motor component and the Eat component, which means that according to the model they are activated whenever the associated psychomotor problem links or eating problem links are significant in a patient-specific network. Similarly, the psychomotor problem and death pre-occupation components both elicit three Guilt-related links. Moreover, the three links between Core, Energy and Cognitive symptoms load on all components and are thus always activated whenever a patient belongs to one or more of the components. Finally, six links involving Eat or Death have zero loadings, implying that the HICLAS model predicts that they are never significant, which means that they never show up in the reconstructed networks.

Component Scores of the Patients

Based on their scores on the three HICLAS components, the 158 patients are divided into eight (2^3) mutually exclusive clusters, as each of the three componants might be present or absent. Figure 6 presents the reconstructed networks of the seven patient clusters that have sig-

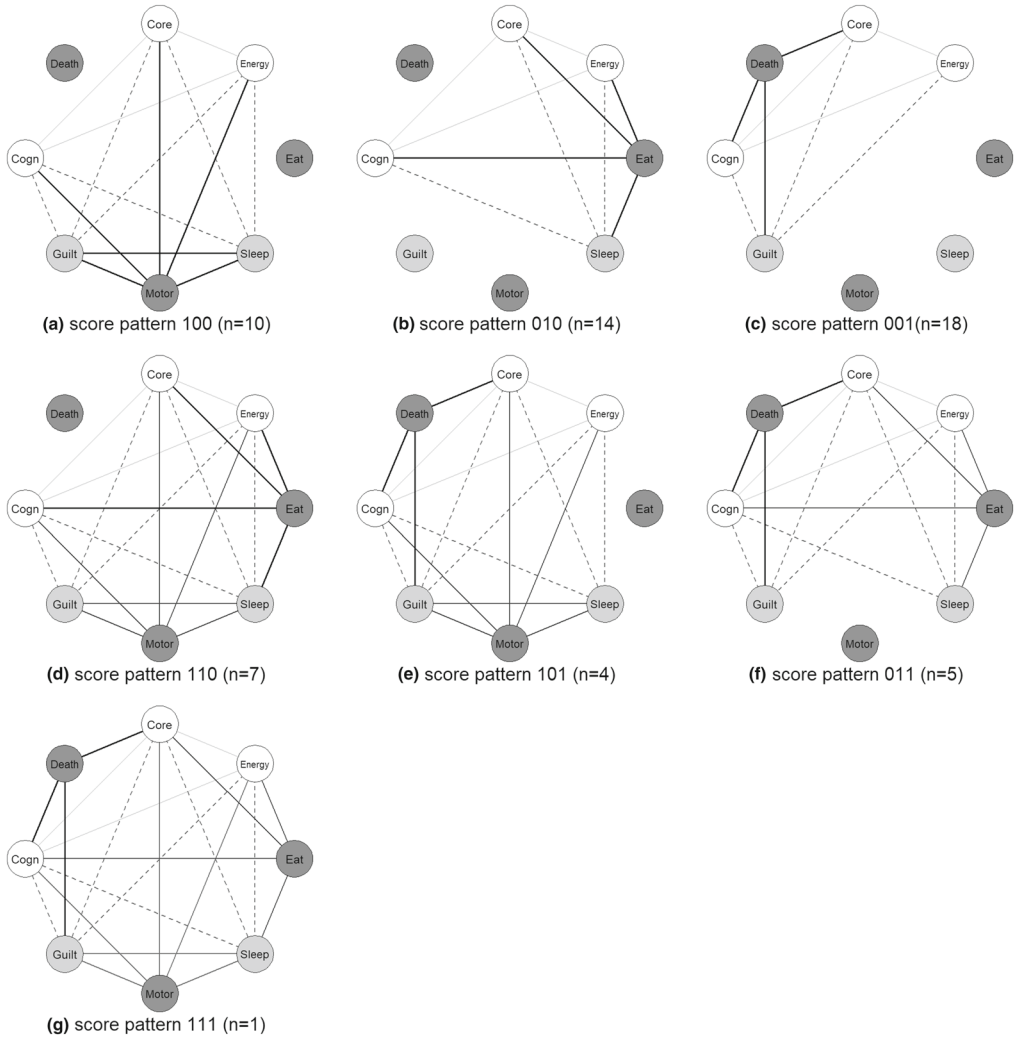


FIGURE 6.
Reconstructed networks of the seven patient clusters according to the HICLAS model with three components. The full black lines represent the component-specific links. The dashed lines indicate links that are shared between two components. The gray lines, finally, represent the links that load on all three components and are thus present for all seven patient clusters that belong to at least one component.

nificant co-occurrence links according to the HICLAS solution. The full black lines represent the component-specific links. These links differentiate maximally between groups of patients. The dashed lines indicate links that are shared between two components. The gray lines, finally, represent the links that load on all three components and are thus activated in all of those seven patient clusters. Therefore, patient clusters 100.[6] (only first component; 10 patients), 010 (only second component; 14 patients) and 001 (only third component; 18 patients) have the sparsest networks that are also quite different, because the associated patients score on one component

[6]Each label consists of three binary scores. When the first score equals one, this indicates that this cluster of patients scores on the first component (clusters 100, 101, 110, 111); when the second score equals one, it scores on the second component (clusters 010, 011, 110, 111); a third score of one implies that it scores on the third component (clusters 001, 011, 101, 111).

only. Cluster 100, for example, scores on component 1 and includes the component specific links (involving Motor), but also the links component 1 shares with component 2 (involving Sleep) or component 3 (involving Guilt) and the links shared with all component (the links between loss of energy, cognitive problems and core symptoms). Patient cluster 111, in contrast, consists of a single patient only, but is characterized by a dense network, because this patient scores on all three components. This network thus includes all component specific links and all shared links. Patient clusters 110 ($n = 7$), 101 ($n = 4$), and 011 ($n = 5$) score on two components and have a network of intermediate complexity. They miss either the Death, the Eat or the Motor links.

The eighth cluster has a score pattern of zeros, indicating that according to the model the 99 persons belonging to this cluster are predicted to not have any significant co-occurrence links. If we look at the ConNEcTs that we obtained for these persons, we see that 57 of them indeed had zero or very few significant links. The other 42 patients had atypical networks, in which the most frequently observed links between Energy, Cognitive and Core symptoms were missing completely ($n = 11$) or partly ($n = 30$); moreover, the links that were significant in the networks of these 42 patients were diverse, which explains why these patients did not form a separate cluster. Indeed, retaining more than three components, which would clearly increase the number of patient clusters, did not considerably lower the number of zero score patients.

HICLAS classified the example patient into patient cluster 001, because most links of component 3 (Death) were present in the associated ConNEcT (see panel c of Fig. 1), while the links of component 1 (Motor) and component 2 (Eat) were absent. Additionally, most of the links that load on multiple components were present. The person-specific network of the example patient, nevertheless, shows two deviations from the reconstructed network of person cluster 001: first, the link Death – Energy is not part of the reconstructed network, but present for the patient. Second, the link Guilt- Energy shows up in the reconstructed network, but is not present in the patient's network.

4. Discussion

This paper had a twofold aim. First, we wanted to introduce the ConNEcT approach as an interesting alternative for studying important psychological concepts like mental disorders (Borsboom, 2017), attitudes (e.g., Dalege et al., 2016), beliefs (e.g., Brandt et al., 2019), and personality (e.g., Cramer et al., 2012). In this paper, we focused on idiographic psychopathological symptom networks. Second, we applied ConNEcT to an extensive data set of longitudinal symptom reports to gain insight into the observed co-occurrence of depression symptoms across time and individual differences therein.

4.1. The ConNEcT Approach

Regarding ConNEcT, the main aim of this paper was to show how ConNEcT can be used to detect person-specific significant co-occurrence links in binary time series data and visualize them. Although other methods and software have been proposed for building such networks, such as mgm and IsingFit, there are some crucial differences between ConNEcT on the one hand and mgm and IsingFit on the other hand: First, ConNEcT calculates, tests and visualizes observed unconditional co-occurrence relations, without making claims about the underlying data-generating model. In contrast, mgm and IsingFit build on a model that specifies the assumed underlying data generating process and focus on conditional relations (e.g., partial correlations) including more than one predictor. Second, mgm and IsingFit do not (yet) take the serial dependence within the time series into account when investigating significance, while ConNEcT comes with a validated permutation-based significance test that allows to account for serial dependencies (Bodner et al. 2020). Third,

ConNEcT can be based on any easy to calculate contingency measures, avoiding identification issues that we encountered with the log-odds ratio.

An R-package for building ConNEcTs from data is under preparation, a beta version is already available on OSF. This package has more functionalities than shown here, in that it also includes the option to lag the variables under consideration. This makes it possible to study sequential dependencies next to co-occurrence (Bodner et al., 2018, 2019; Van keer et al., 2019). Moreover, next to the Jaccard measure, other association measures are available, such as Cohen's Kappa and the proportion of agreement. Moreover, it is perfectly possible to include and focus on the log-odds ratio, although one should then keep the issues in mind that we pointed out in the section on model-based approaches.

4.2. Application to the Hosenfeld et al. Data

Regarding our illustrative application, we built ConNEcTs of eight depression symptoms for 158 patients, revealing strong individual differences. To capture the main similarities and differences in these networks, we clustered them using HICLAS (De Boeck & Rosenberg, 1988). The obtained solution showed that some links only occur for specific patient groups, whereas other links are more general. Specifically, we found three components, each being uniquely characterized by links centered around one symptom (preoccupation with death, eating and psychomotor problems). These links thus only show up in specific patient subgroups. On the other hand, we also found some links that are more universal and thus differentiate less between patients. These links focus on the pairwise co-occurrence of cognitive symptoms, core symptoms and loss of energy. In between, we find two groups of links, one pertaining to links with sleeping problems and one to links with feelings of guilt, that are shared by some patient groups, but not by all. These results are in line with other findings that have been reported in the literature. For instance, the differentiation between melancholic and atypic depression parallels the components that we find in our data: while atypic depression is characterized by sleeping and eating problems (combined with weight gain), melancholic depression seems to be characterized by symptoms like psychomotor problems and excessive feelings of guilt (Lamers et al., 2010). Similarly, Wardenaar et al. (2015) distinguish somatic and cognitive depression, whereby the sleep links are counted by the somatic symptoms and guilt links by the cognitive symptoms.

At first sight, it might be disconcerting that 99 patients did not score on any of the components, implying that their networks were empty according to the HICLAS solution. However, out of these 99 patients, 41 patients had two significant links at most, which is quite in line with the HICLAS reconstruction. The other 48 patients had very idiosyncratic networks that hardly resembled one another, explaining why some of them did not form a separate cluster. The latter finding links up with earlier findings revealing huge heterogeneity in how symptoms are combined in depressed patients (e.g., Fried et al., 2016). The first finding might be unexpected, given that Borsboom's network theory of mental disorders (Borsboom, 2017) seems to suggest that connectivity is a necessary ingredient for depression. However, this finding might be due to the treatment that these patients received at the start of the study, which might have been successful in breaking some of this connectivity. From a dynamic systems point of view, it is indeed likely that recovery is preceded by a phase of weakening the depressed patterns (Olthof et al., 2020; van de Leemput et al., 2014)

The approach applied in this paper consisted of two steps. In each step, we had to make some decisions that can obviously influence the results to some extent. A very important decision in the first step pertains to the applied significance level, as this immediately determines how many links are significant. Given our idiographic focus, a large number of symptom links (i.e., 28) had to be tested for each considered individual separately. This large number may raise concerns regarding multiple testing and false positives. While we did not implement a classical correction for

multiple testing (e.g., Benjamini & Hochberg, 1995), we investigated the impact of this decision, by applying a multi verse approach (Steegen et al., 2016) and inspecting other significance levels than .05. Specifically, we also considered .001, .01, .04, .05, .06, .1. Subsequently, we compared the obtained HICLAS solutions (details of the results can be consulted in the supplementary material). The overall patterns revealed by HICLAS seem rather robust and the changes predictable: higher significance levels lead to more links being significant, implying that more links show up in the reconstructed networks that HICLAS yields. The additional links are in most cases added to the most likely component. When changing the alpha level from .05 to .06, for instance, two Death-related links are included and added to the Death component. Of course, when handling very strict significance levels (e.g., .001) most reconstructed HICLAS networks are empty. In the second step, the most important decision pertains to the clustering approach used. Here, we opted to use HICLAS as this approach was specifically developed for binary data, but of course other choices such as latent class analysis (Vermunt & Magidson, 2002) are possible and defendable.

4.3. Limitations and Future Directions

Although we refrained from correction for multiple testing for now (see higher), we do recognize that this might have led to false positives. For future research, it might therefore be useful to look into several possibilities to reduce the number of tests that have to be performed. A variable selection or dimension reduction approach could be applied on the variables, to reduce or combine the number of variables under study, effectively lowering the number of tests (Bulteel, Tuerlinckx, et al., 2018). On the person side, one could try to group data from similar persons (Brusco et al., 2019), after which significance could be tested per group rather than per individual. One could even check whether the observed co-occurrence patterns are indicative of one data generating model that holds for everyone, but from which specific subsets of connections are dropped because specific symptoms do not occur for particular persons. If this would be the case, the tests could maybe be run across persons, effectively decreasing the number of tests.

Second, we focused on concurrent inter-dependence in this study. It may also be interesting to investigate sequential dependencies between symptoms: Does the presence of one symptom in a specific week, consistently activate another symptom the next week. To this end, one of the symptoms can simply be lagged, after which the ConNEcT procedure can be applied again. Answering more intricate conditional interdependence questions is not possible for the moment, however. For instance, one may want to know how strong the concurrent relationship is, while controlling for lagged relations on top of serial dependency, or vice versa.

Third, we now only focused on the significance of a specific link or edge, using a non-parametric permutation test. For future research, it would be useful to also obtain an idea about the standard errors of the computed contingency measures, to assess the precision of the estimates. (Blocked) leave one-out procedures (Racine, 2000) might be an interesting direction here that would retain most of the temporal dependence in the data.

Fourth, ConNEcT is very flexible and allows researchers to plug in a wide variety of contingency measures. Here, for reasons of readability, we limited ourselves to the Jaccard index. We chose this index for a number of substantive reasons (see *Step 2: Quantification of co-occurrence*). Thoroughly scrutinizing alternative measures fell out of the scope of this paper. Nevertheless, investigating the advantages and disadvantages of different contingency measures in an extensive simulation study is a necessary next step to provide researchers with recommendations about which measure to use in which context and for which research question.

5. Conclusion

To summarize, we presented ConNEcT, a novel network approach for investigating the observed co-occurrence of binary variables over time. Reanalyzing data from an extensive study on depression patients, we revealed strong heterogeneity in observed symptom co-occurrence, but also some meaningful patient clusters.

Acknowledgments

The research leading to the results reported in this paper was sponsored in part by research grants from the Fund for Scientific Research-Flanders (FWO, Project No. G066316N awarded to Eva Ceulemans and Francis Tuerlinckx) and by the Research Council of KU Leuven (C14/19/054 and PDM/20/062). The raw data analyzed during the current study are available from the corresponding author on reasonable request. HICLAS software can be requested by emailing eva.ceulemans@kuleuven.be.. The beta version of the R package ConNEcT (0.6.3) can be downloaded from OSF at https://osf.io/rfxhc/.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix

Retrieving model-based Networks with ConNEcT

ConNEcT is a widely applicable method for capturing and testing contingency of binary variables across time. It is not based on a particular model specification. Therefore, we expect that it is able to find associations irrespective of which underlying model generated these associations. To demonstrate this, we will apply ConNEcT to binary data that have been generated by published syntaxes. These syntaxes pertain to a latent variable model (Golino & Epskamp, 2017) and the Curie-Weiss network and sparse network from the article of Epskamp et al. (2017). Both models are cross-sectional, as we did not readily find scripts for generating longitudinal binary data.

Application 1: Retrieving the Associations Implied by a Latent Variable Model

As a first example, we use the syntax of Golino and Epskamp (2017), who generate binary data from a latent variable model. These authors make use of exploratory graph analysis, as a specific form of network psychometrics that looks for clusters of highly connected variables, to retrieve the latent variable structure. They expect that each latent variable will correspond to a different cluster. Therefore, we will investigate whether applying ConNEcT to such data will also reveal such clusters of highly connected variables.

Design of the Simulation Study of Golino and Epskamp
Number of factors (latent variables): 2, 4
Number of variables per factor: 5,10
Number of observations: 100
Correlation between factors: 0, 0.3, 0.5, 0.7
For details on the data generation procedure, we refer to the original paper. We generated 100 data sets for each of the 16 settings and analyzed each data set with the ConNEcT approach. Note that the data contain no serial dependence as the observations within each dataset were sampled independently.

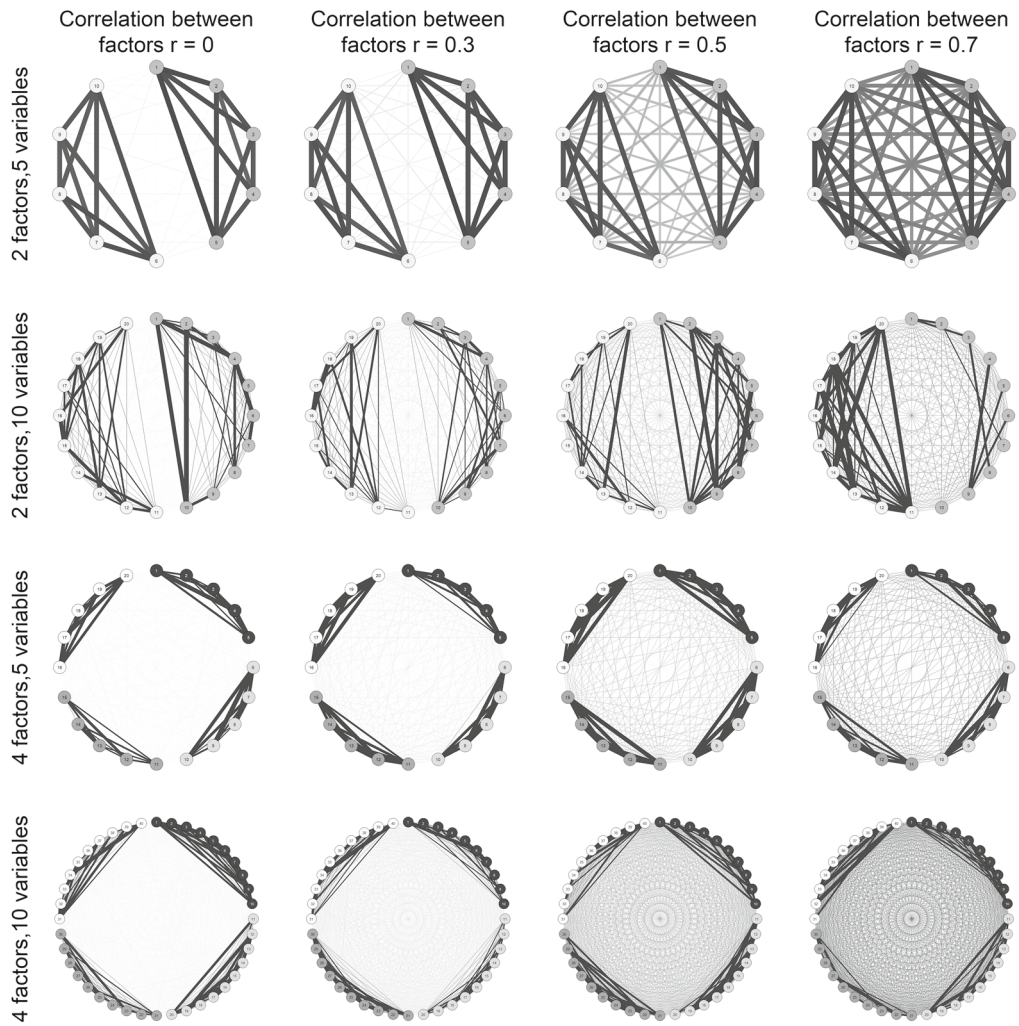


FIGURE 7.
Retrieved ConNEcTs for 16 different latent variable models that imply two factors (upper two rows) or four factors (lower two rows), with 5 (rows 1 and 3) or 10 (row 2 and 4) variables per factor. The intercorrelations among the factors amount to 0 (first column), 0.3 (second column), 0.5 (third column) and 0.7 (last column). Different node colors indicate generation from the same factor. The number of significant links is summarized over all 100 simulated data sets per simulation setting. Dark lines indicate that a link is significant in almost all data sets, while less saturated or thinner links show links that are significant in few data sets only.

Settings ConNEcT Significance Test The scores on each variable were divided into ten segments, which were permuted 100 times. For each combination of two variables, we considered all possible pairs of the corresponding permuted variables, resulting in a sampling distribution of 10,000 Jaccard values.

Results We computed for each variable pair how often its association was significant across the 100 simulated data sets. These results are visualized in the networks in Fig. 7, where the width and saturation of the network edges represent how often the associated link was significant. Dark lines therefore indicate a link that is significant in almost all data sets, while less saturated or thinner links show links that are significant in a few data sets only. The networks reveal that the ConNEcT

approach indeed retrieves the latent variable models, in that the variables that load strongly on the same underlying factor are densely connected. If the factors are not correlated, we hardly find any links between different clusters (first column). However, the higher the correlations between the factors, the more between-cluster links are found, as could be expected.

Application 2: Retrieving the Association Structure of a Curie Weiss Network and a Sparse Network

In this section, we use the Curie–Weiss and sparse networks that were studied by Epskamp, Kruis and Marsman (2017 see Fig. 1 of their paper). For both networks, we generate 100 data sets, each consisting of 10 variables and 1000 independent observations, and test the significance of each pairwise link with the ConNEcT approach. We used the same test settings as in Study 1.1. The number of significant links across the 100 data sets is shown in Fig. 8, revealing that we indeed retrieve the underlying network structures.

Comparison of Model-based and ConNEcT-Based Testing Results

To demonstrate that not accounting for serial dependence in the data indeed leads to type 1 errors, we generated data using the simulation syntaxes of Bodner et al. (2020). We twice simulated 100 datasets, that each contain ten independent 'time series' of 500 observations. In one setting, the data contain no serial dependence; in the other setting they do. Subsequently, we investigate how often each of the $10^*9/2 = 45$ links between the variables is considered significant by the IsingFit (van Borkulo et al., 2014) and mgm (Haslbeck & Waldorp, 2020) approaches. As the variables are generated independently, one might naively expect a type 1 error rate (i.e., percentage of false positives) of around 5%. However, we expect that the type 1 error rate will only be reasonable for the data without serial dependence and will be too high for the data with serial dependence. We will compare the results to those of the ConNEcT approach.

Data Generation

The data were generated by drawing from one or several Bernoulli distributions:
data were generated by drawing from one or several Bernoulli distributions:

- For the data without serial dependence, all observations were drawn form a Bernoulli distribution with probability $p_1 = .2$.
- For the data with serial dependence, the first observation was again drawn from a Bernoulli distribution with probability $p_1 = .2$; the following observations were either drawn from a Bernoulli distribution with $p_{1|1} = .9$, if the previous observation had been a 1 or from a distribution with $p_{1|0} = .025$ (this value was computed based on the chosen values for p_1 and $p_{1|1}$: $p_{1|0} \sim (p_1\text{-}p_1{}^*p_{1|1})/(1\text{-}p_1)$; see Bodner et al. 2020).

We decided to take data sets of 500 observations, because longer time series will match the chosen prevalence and serial dependence level more exactly. This is especially important, because for both IsingFit and mgm to run smoothly, we need to make sure that there are enough 1s observed (while the IsingFit gives a warning when less than eight 1s are included, the mgm does not give a result if variables contain less than three 1s).

Analysis and Results

For each method (ConNEcT, IsingFig, mgm) and each data type (with or without serial dependence, we summed the significance results across the 100 simulated data sets. As expected, all three methods perform well, when the data contain no serial dependence:

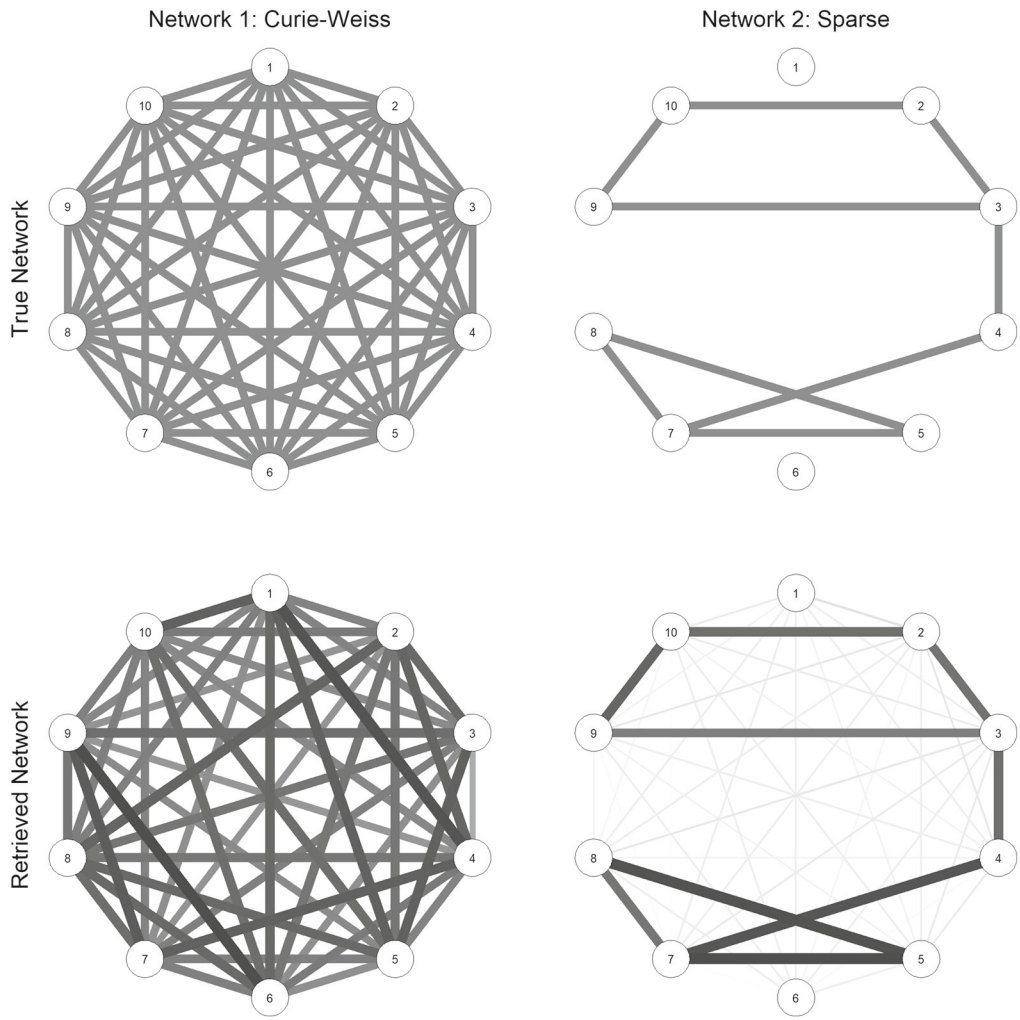


FIGURE 8.
Retrieving **a** a Curie–Weiss network and **b** a sparse network; 100 data sets were generated each consisting of 10 variables with n = 1000 observations and tested with the ConNEcT approach. Dark lines indicate link that are significant in almost all data sets, while less saturated or thinner links show links that are significant only in few data sets.

- For IsingFit, the type 1 error rate per link never exceeds the nominal level of 5%. The false positive rate ranges between 0 and 0.02 (mean: 0.003), which is actually too small.
- For mgm, the type 1 error rate ranges between .01 and .09 (mean: 0.048), which is quite satisfactory.
- For ConNEcT, we find a type 1 error rate of 0 to 0.07 (mean: .033) (Fig. 9).

However, the results for the data that contain serial dependence clearly show that taking into account serial dependence matters and that using a method that does not do this yields a too high type 1 error rate.

- For the IsingFit, the type 1 error rate ranges between .36 and .62 (mean: .46)
- For mgm , the type 1 error ranges between .54 and .75 (mean: .59)
- For ConNEcT, the type 1 error rate ranges between 0 and 0.12 (mean: 0.05)

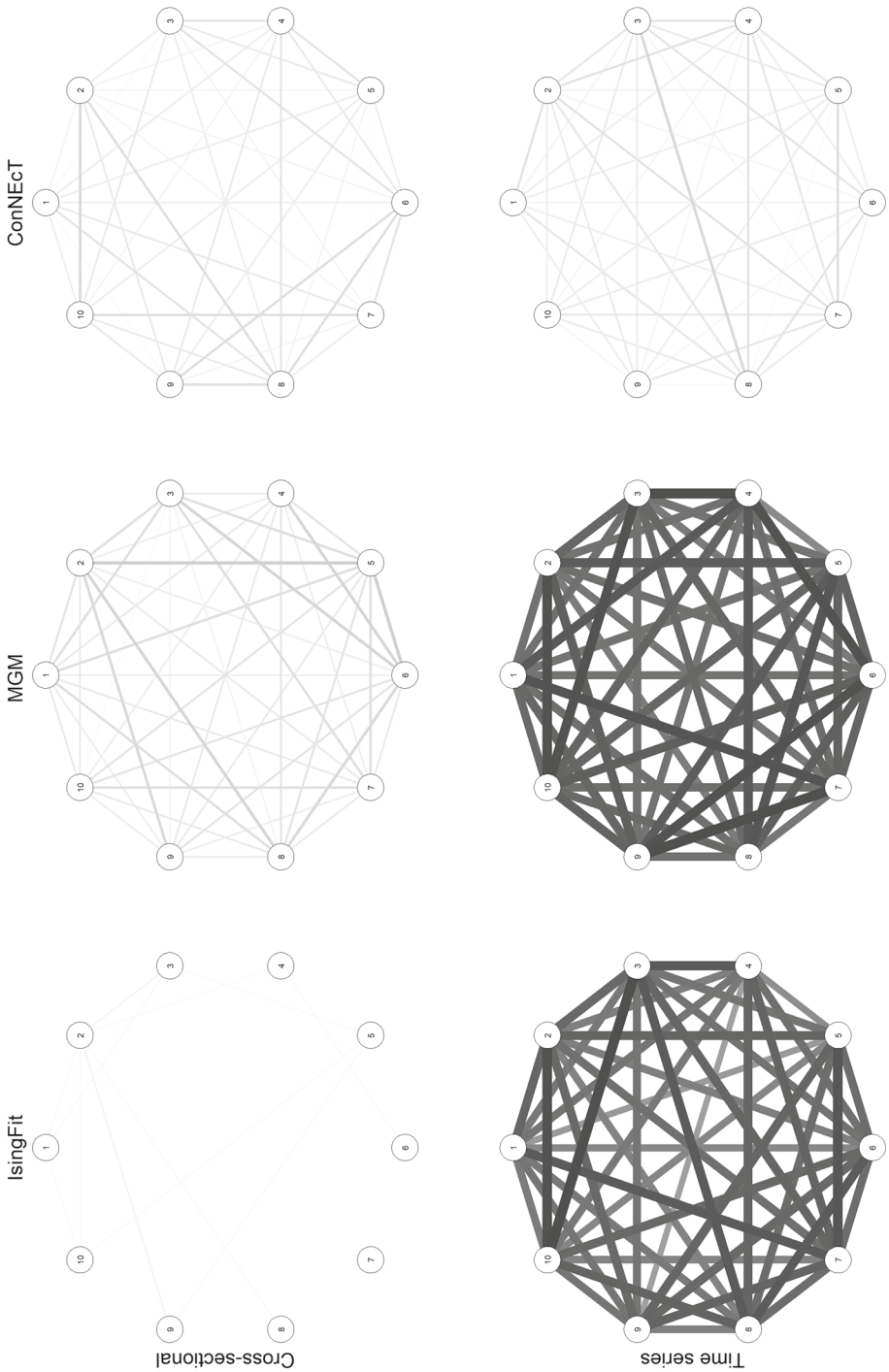


FIGURE 9.

Investigating proportion false positives for independent data 1) without serial dependence (first row) and 2) with serial dependence (second row) with **a** IsingFit, **b** mgm and **c** ConNEcT. 100 data sets of 10 independent variables (length n=500) were drawn from 1) Bern(0.2) for the cross-sectional data or 2) time point per time point from a Bernoulli distribution with probability $p(X_t = 1|X_{t-1} = 1) = .9$ or $p(X_t = 1|X_{t-1} = 0) = .025$ for the time series data.

References

Albert, A., & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, *71*(1), 1–10. https://doi.org/10.1093/biomet/71.1.1.

Bakeman, R., McArthur, D., & Quera, V. (1996). Detecting group differences in sequential association using sampled permutations: Log odds, kappa, and phi compared. *Behavior Research Methods, Instruments, and Computers*, *28*(3), 446–457. https://doi.org/10.3758/BF03200524.

Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge: Cambridge University Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bodner, N., Bosmans, G., Sannen, J., Verhees, M., & Ceulemans, E. (2019). Unraveling middle childhood attachment-related behavior sequences using a micro-coding approach. *Plos One*, *14*(10), e0224372. https://doi.org/10.1371/journal.pone.0224372.

Bodner, N., Kuppens, P., Allen, N. B., Sheeber, L. B., & Ceulemans, E. (2018). Affective family interactions and their associations with adolescent depression: A dynamic network approach. *Development and Psychopathology*, *30*(4), 1459–1473. https://doi.org/10.1017/S0954579417001699.

Bodner, N., Tuerlinckx, F., Bosmans, G., & Ceulemans, E. (2020). Accounting for auto-dependency in binary dyadic time series data: A comparison of model- and permutation-based approaches for testing pairwise associations. *British Journal of Mathematical and Statistical Psychology*. https://doi.org/10.1111/bmsp.12222

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*(9), 1089–1108. https://doi.org/10.1002/jclp.20503.

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375.

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*(1), 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608.

Boschloo, L., van Borkulo, C. D., Rhemtulla, M., Keyes, K. M., Borsboom, D., & Schoevers, R. A. (2015). The network structure of symptoms of the diagnostic and statistical manual of mental disorders. *Plos One*, *10*(9), e0137621. https://doi.org/10.1371/journal.pone.0137621.

Bosley, H. G., Sandel, D. B., & Fisher, A. J. (2020). Idiographic dynamics of positive affect in gad: Modeling emotion regulation at the person level. *European Journal of Psychological Assessment*, *36*(3), 500–509. https://doi.org/10.1027/1015-5759/a000580.

Brandt, M. J., Sibley, C. G., & Osborne, D. (2019). What is central to political belief system networks? *Personality and Social Psychology Bulletin*, *45*(9), 1352–1364. https://doi.org/10.1177/0146167218824354.

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., et al. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS ONE*, *8*(4), e60188. https://doi.org/10.1371/journal.pone.0060188.

Brusco, M. J., Steinley, D., Hoffman, M., Davis-Stober, C., & Wasserman, S. (2019). On Ising models and algorithms for the construction of symptom networks in psychopathological research. *Psychological Methods*, *24*(6), 735–753. https://doi.org/10.1037/met0000207.

Bulteel, K., Mestdagh, M., Tuerlinckx, F., & Ceulemans, E. (2018). VAR (1) based models do not always outpredict AR(1) models in typical psychological applications. *Psychological Methods*, *23*(4), 740–756. https://doi.org/10.1037/met0000178.

Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2018). Improved insight into and prediction of network dynamics by combining VAR and dimension reduction. *Multivariate Behavioral Research*, *53*(6), 853–875. https://doi.org/10.1080/00273171.2018.1516540.

Ceulemans, E., & Van Mechelen, I. (2005). Hierarchical classes models for three-way three-mode binary data: Interrelations and model selection. *Psychometrika*, *70*(3), 461–480. https://doi.org/10.1007/s11336-003-1067-3.

Ceulemans, E., Van Mechelen, I., & Leenen, I. (2007). The local minima problem in hierarchical classes analysis: An evaluation of a simulated annealing algorithm and various multistart procedures. *Psychometrika*, *72*(3), 377–391. https://doi.org/10.1007/s11336-007-9000-9.

Ceulemans, E., & Kiers, H. A. L. (2006). Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. *British Journal of Mathematical and Statistical Psychology*, *59*(1), 133–150. https://doi.org/10.1348/000711005X64817.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104.

Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., et al. (2016). Major depression as a complex dynamic system. *Plos One*, *11*(12), e0167490. https://doi.org/10.1371/journal.pone.0167490.

Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., et al. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, *26*(4), 414–431. https://doi.org/10.1002/per.1866.

Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, *33*(2–3), 137–150. https://doi.org/10.1017/S0140525X09991567.

Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. J. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychological Review*, *123*(1), 2–22. https://doi.org/10.1037/a0039802.

De Boeck, P., & Rosenberg, S. (1988). Hierarchical classes: Model and data analysis. *Psychometrika*, *53*(3), 361–381. https://doi.org/10.1007/BF02294218.

de Vos, S., Wardenaar, K. J., Bos, E. H., Wit, E. C., Bouwmans, M. E. J., & de Jonge, P. (2017). An investigation of emotion dynamics in major depressive disorder patients and healthy persons using sparse longitudinal networks. *Plos One*, *12*(6), e0178586. https://doi.org/10.1371/journal.pone.0178586.

Edgington, E. S., & Onghena, P. (2007). *Randomization tests* (4th ed.). Florida: Chapman & Hall/CRC.

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1.

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 1–18.

Epskamp, S., Kruis, J., & Marsman, M. (2017). Estimating psychopathological networks: Be careful what you wish for. *Plos One*, *12*(6), e0179891. https://doi.org/10.1371/journal.pone.0179891.

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*(1), 27–38. https://doi.org/10.1093/biomet/80.1.27.

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of Abnormal Psychology*, *126*(8), 1044–1056. https://doi.org/10.1037/abn0000311.

Fisher, R. A. (1949). *The design of experiments* (5th ed.). Oliver & Boyd.

Fried, E. I., van Borkulo, C. D., Cramer, A. O. J., Boschloo, L., Schoevers, R. A., & Borsboom, D. (2017). Mental disorders as networks of problems: A review of recent insights. *Social Psychiatry and Psychiatric Epidemiology*, *52*(1), 1–10. https://doi.org/10.1007/s00127-016-1319-z.

Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders*, *189*, 314–320. https://doi.org/10.1016/j.jad.2015.09.005.

Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders*, *172*, 96–102. https://doi.org/10.1016/j.jad.2014.10.010.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191.

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), e0174035. https://doi.org/10.1371/journal.pone.0174035.

Good, P. (2000). *Permutation Tests*. Springer New York. https://doi.org/10.1007/978-1-4757-3235-1

Hamaker, E., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study. *Multivariate Behavioral Research*, *53*(6), 820–841. https://doi.org/10.1080/00273171.2018.1446819.

Hamming, R. W. (1950). Error detecting and error correcting codes. *The Bell System Technical Journal*, *29*(2), 147–160. https://doi.org/10.1002/j.1538-7305.1950.tb00463.x.

Haslbeck, J. M. B., & Waldorp, L. J. (2020). mgm: Estimating time-varying mixed graphical models in high-dimensional data. ArXiv:1510.06871 [Stat].

Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, *21*(16), 2409–2419. https://doi.org/10.1002/sim.1047.

Holloway, E. L., Wampold, B. E., & Nelson, M. L. (1990). Use of a paradoxical intervention with a couple: An interactional analysis. *Journal of Family Psychology*, *3*(4), 385–402. https://doi.org/10.1037/h0080552.

Hosenfeld, B., Bos, E. H., Wardenaar, K. J., Conradi, H. J., van der Maas, H. L. J., Visser, I., et al. (2015). Major depressive disorder as a nonlinear dynamic system: Bimodality in the frequency distribution of depressive symptoms over time. *BMC Psychiatry*, *15*(1), 222. https://doi.org/10.1186/s12888-015-0596-5.

Isvoranu, A.-M., Borsboom, D., van Os, J., & Guloksuz, S. (2016). A network approach to environmental impact in psychotic disorder: Brief theoretical framework. *Schizophrenia Bulletin*, *42*(4), 870–873. https://doi.org/10.1093/schbul/sbw049.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de La Société Vaudoise Des Sciences Naturelles*, *37*, 547–579.

Jaccard, P. (1912). The distribution of the flora in the Alpine zone. *New Phytologist*, *11*(2), 37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x.

Lamers, F., de Jonge, P., Nolen, W. A., Smit, J. H., Zitman, F. G., Beekman, A. T. F., et al. (2010). Identifying depressive subtypes in a large Cohort study: Results from the Netherlands study of depression and anxiety (NESDA). *The Journal of Clinical Psychiatry*, *71*(12), 1582–1589. https://doi.org/10.4088/JCP.09m05398blu.

Mansournia, M. A., Geroldinger, A., Greenland, S., & Heinze, G. (2018). Separation in logistic regression: Causes, consequences, and control. *American Journal of Epidemiology*, *187*(4), 864–870. https://doi.org/10.1093/aje/kwx299.

Marsman, M., Waldorp, L., & Borsboom, D. (2019). *Towards an encompassing theory of network models* [Preprint]. https://doi.org/10.31234/osf.io/n98qt

Moulder, R. G., Boker, S. M., Ramseyer, F., & Tschacher, W. (2018). Determining synchrony between behavioral time series: An application of surrogate data generation for establishing falsifiable null-hypotheses. *Psychological Methods*, *23*(4), 757–773. https://doi.org/10.1037/met0000172.

Olthof, M., Hasselman, F., Strunk, G., van Rooij, M., Aas, B., Helmich, M. A., et al. (2020). Critical fluctuations as an early-warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science*, *8*(1), 25–35. https://doi.org/10.1177/2167702619865969.

Racine, J. (2000). Consistent cross-validatory model-selection for dependent data: Hv-block cross-validation. *Journal of Econometrics*, *99*(1), 39–61. https://doi.org/10.1016/S0304-4076(00)00030-0.

Salvatore, S., Rand, K. D., Grytten, I., Ferkingstad, E., Domanska, D., Holden, L., Gheorghe, M., Mathelier, A., Glad, I., & Sandve, G. K. (2019). Beware the Jaccard: The choice of metric is important and non-trivial in genomic colocalisation analysis. *BioRxiv*, 479253. https://doi.org/10.1101/479253.

Scott, W. A. (1955). Reliability of content analysis: The case of nominal scale coding. *The Public Opinion Quarterly*, *19*(3), 321–325. https://www.jstor.org/stable/2746450.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637.

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

van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., et al. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences, 111*(1), 87–92. https://doi.org/10.1073/pnas.1312114110.

Van Keer, I., Ceulemans, E., Bodner, N., Vandesande, S., Van Leeuwen, K., & Maes, B. (2019). Parent-child interaction: A micro-level sequential approach in children with a significant cognitive and motor developmental delay. *Research in Developmental Disabilities*, *85*, 172–186. https://doi.org/10.1016/j.ridd.2018.11.008.

Van Mechelen, I., De Boeck, P., & Rosenberg, S. (1995). The conjunctive model of hierarchical classes. *Psychometrika*, *60*(4), 505–521. https://doi.org/10.1007/BF02294326.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. Hagenaars & A. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge: Cambridge University Press.

Wardenaar, K. J., Monden, R., Conradi, H. J., & de Jonge, P. (2015). Symptom-specific course trajectories and their determinants in primary care patients with Major Depressive Disorder: Evidence for two etiologically distinct prototypes. *Journal of Affective Disorders*, *179*, 38–46. https://doi.org/10.1016/j.jad.2015.03.029.

Wilderjans, T. F., Ceulemans, E., & Meers, K. (2013). CHull: A generic convex-hull-based model selection method. *Behavior Research Methods*, *45*(1), 1–15. https://doi.org/10.3758/s13428-012-0238-5.

Wright, A. G. C., & Woods, W. C. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology*, *16*(1), 49–74. https://doi.org/10.1146/annurev-clinpsy-102419-125032.

Manuscript Received: 13 MAR 2020
Final Version Received: 8 APR 2021
Accepted: 16 APR 2021
Published Online Date: 1 JUN 2021