

Routledge
Taylor & Francis Group

Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: http://www.tandfonline.com/loi/hmbr20

Using Raw VAR Regression Coefficients to Build Networks can be Misleading

Kirsten Bulteel, Francis Tuerlinckx, Annette Brose & Eva Ceulemans

To cite this article: Kirsten Bulteel, Francis Tuerlinckx, Annette Brose & Eva Ceulemans (2016) Using Raw VAR Regression Coefficients to Build Networks can be Misleading, Multivariate Behavioral Research, 51:2-3, 330-344, DOI: 10.1080/00273171.2016.1150151

To link to this article: http://dx.doi.org/10.1080/00273171.2016.1150151

Published online: 30 Mar 2016.

Submit your article to this journal

Article views: 98

View related articles

View Crossmark data

Citing articles: 2 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=hmbr20

Download by: [KU Leuven University Library] **Date:** 26 October 2016, At: 05:26

MULTIVARIATE BEHAVIORAL RESEARCH
2016, VOL. 51, NOS. 2–3, 330–344
http://dx.doi.org/10.1080/00273171.2016.1150151

Routledge
Taylor & Francis Group

Using Raw VAR Regression Coefficients to Build Networks can be Misleading

Kirsten Bulteel[a], Francis Tuerlinckx[a], Annette Brose[b], and Eva Ceulemans[a]

[a]Faculty of Psychology and Educational Sciences, KU Leuven; [b]Institute for Psychology, Humboldt University of Berlin

ABSTRACT

Many questions in the behavioral sciences focus on the causal interplay of a number of variables across time. To reveal the dynamic relations between the variables, their (auto- or cross-) regressive effects across time may be inspected by fitting a lag-one vector autoregressive, or VAR(1), model and visualizing the resulting regression coefficients as the edges of a weighted directed network. Usually, the raw VAR(1) regression coefficients are drawn, but we argue that this may yield misleading network figures and characteristics because of two problems. First, the raw regression coefficients are sensitive to scale and variance differences among the variables and therefore may lack comparability, which is needed if one wants to calculate, for example, centrality measures. Second, they only represent the unique direct effects of the variables, which may give a distorted picture when variables correlate strongly. To deal with these problems, we propose to use other VAR(1)-based measures as edges. Specifically, to solve the comparability issue, the standardized VAR(1) regression coefficients can be displayed. Furthermore, relative importance metrics can be computed to include direct as well as shared and indirect effects into the network.

KEYWORDS

Network modeling; regression analysis; relative importance; standardization; vector autoregressive modeling

Many questions in the behavioral sciences focus on the causal interplay of a number of variables within individuals across time. Consider for example the work of Borsboom and Cramer (2013) in the field of psychopathology. These authors emphasize the importance of studying the causal relations between the symptoms of a disorder across time (e.g., insomnia, concentration problems, and fatigue in the case of depression) to understand why and how people develop and maintain a particular disorder. In addition, tracing which symptoms have the largest causal effect on the other symptoms is useful to identify targets for interventions. As another example, Pe and colleagues examined whether the experience of specific emotions increases or decreases the intensity of other emotions at the next timepoint (i.e., augmentation or blunting, respectively; Pe & Kuppens, 2012) and speculate that the overall strength of these lagged relations is related to depression (Pe et al., 2015).

Such research questions are currently often answered by applying regression-based methods, which allow inspecting how variables affect themselves (i.e., autoregressive effects) and each other (i.e., cross-regressive effects) across time. In this article, we will focus on the most often used regression-based method so far—that is, the reduced-form vector autoregressive modeling of order 1 (VAR(1); e.g., Bos, Hoenders, & de Jonghe, 2012; Bringmann et al., 2013; Pe et al., 2015; Schmitz & Skinner, 1993; Snippe et al., 2015; van Gils et al., 2014; Wichers, 2014). In a VAR(1) model, each variable is regressed on all variables at the previous timepoint (including itself; Hamilton, 1994, Chapter 11; Lütkepohl, 2005, Chapter 2), whereas possible contemporaneous relations between the variables are captured by allowing for correlations among the respective residuals rather than by including separate regression weights (hence, the term *reduced form*). VAR(1) is a fully exploratory approach (i.e., all lag-one relations are considered), and the issues we will raise are therefore most apparent for this technique. However, they are also relevant when applying other regression-based techniques as we will elaborate in the Discussion section.

To gain a deeper and more comprehensive insight into the causal dynamics, the VAR(1) auto- and cross-regressive effects can be displayed in a weighted directed network figure. Such figures are labeled dynamic networks (Bringmann, Lemmens, Huibers, Borsboom, & Tuerlinckx, 2015; Wichers, 2014). In these networks, the individual's variables are the nodes or vertices, and the regression weights constitute the edges (i.e., the arrows between the nodes), which visualize how the variables affect each other over time. Note that similar graphs are regularly drawn when using structural equation models. Next to this visual overview of all dynamic relations, the network approach offers several network characteristics that can be computed on the basis of the edges (see

CONTACT Kirsten Bulteel kirsten.bulteel@ppw.kuleuven.be Faculty of Psychology and Educational Sciences, KU Leuven, Tiensestraat 102, Box 3713, 3000 Leuven, Belgium

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hmbr.

© 2016 Taylor & Francis Group, LLC

Newman, 2010, Chapter 7). For example, centrality characteristics allow identifying which variables strongly incite multiple other variables when activated, which is important information when designing interventions.

The aim of this article is to argue that inspecting only the raw VAR(1) edges, as is commonly done in network applications, could give rise to misinterpretations regarding the causal dynamics. We address two particular sources of misinterpretation: (1) These edges might lack comparability, and (2) they reflect only the unique direct effects among the variables. Regarding the issue of comparability, it is tempting to directly compare the thickness of the edges to detect strong variable connections. However, as is known from regression analysis, scale or variance differences among the variables render such a direct comparison meaningless as these differences are reflected in the regression coefficients and thus in the edges. As a consequence, computing network characteristics on the basis of these edges also makes no sense. In this article, we will limit the discussion to centrality metrics, although the arguments raised apply to all other edge-based network characteristics.

With respect to the type of effects captured, if two nodes are only weakly connected (i.e., weak unique direct effects), researchers can be inclined to infer that two nodes hardly influence one another. This conclusion may be incorrect as the edges do not reflect possible shared or indirect effects among nodes across time. Specifically, the estimated regression effects only reflect Granger causal relations, which can be defined as follows: If a variable A contains unique information to predict the future state of variable B (thus, on top of the current state of variable B and of other variables), then A Granger causes B (Granger, 1969; Lütkepohl, 2005). This implies that shared effects of variables A and B on a third variable, C (both variables have information in common that allows to predict the future state of variable C on top of the current state of variable C and of other variables), or the indirect causal effects of A on C via B (variable A influences variable C through variable B) are not captured in the multivariate regression coefficients, whereas such effects are often also interesting in practice.

Neither comparability nor unique effects has been considered in the literature on VAR models. This can probably be explained by noting that VAR is mostly used for forecasting, implying that proper interpretation of the obtained regression weights is less important. Yet these issues have been the subject of intensive debate in the context of traditional regression analysis. The use of standardized regression coefficients has been advocated to establish comparability, and measures of the relative importance of the predictors have been proposed to take shared effects into account (see, e.g., Grömping, 2007; Johnson & LeBreton, 2004). In this article, we draw on this strand of literature to propose two alternative ways of constructing a VAR(1)-based network, yielding a standardized network and a relative importance one, and illustrate that they may lead to quite different conclusions. Which alternative should be selected depends on the research question at hand.

Table 1. Means and variances of the eleven depression-related symptoms and rating scale used.

Symptom	Scale (end points)	Mean	Variance
Physical symptoms	4-point (0–3)	0.51	0.12
Rumination	8-point (0–7)	4.61	0.57
Feeling guilty	8-point (0–7)	0.55	0.49
Feeling unhappy	8-point (0–7)	− 2.72	0.59
Feeling downhearted	8-point (0–7)	2.79	1.64
Loss of activation	8-point (0–7)	− 2.36	0.66
Loss of interest	8-point (0–7)	− 2.40	0.31
Poor sleep quality	8-point (0–7)	− 2.70	0.23
Cognitive problems	4-point (0–3)	1.41	0.71
Loss of energy	4-point (0–3)	1.54	0.90
Restlessness	4-point (0–3)	1.14	0.97

The remainder of the article is organized as follows. In the following section, we introduce a data set on depression-related symptoms that will be used as an example throughout the article. In the third section, the network approach for time series data of a single individual is presented and discussed. This section ends with an elaboration of the two identified problems, the comparability and the unique-effect issue. The fourth and fifth sections describe alternative edge metrics to overcome these problems. In the fourth section, we discuss the use of standardized regression coefficients. In the fifth section, a relative importance metric covering the unique direct as well as the shared effects is presented. Some concluding remarks and directions for future research will be addressed in the final section of the article.

The example

The data that will be used as an illustrative example throughout the article stem from a 21-year old woman who participated in the COGITO study (Schmiedek, Lövdén, & Lindenberger, 2009). We selected this particular individual because her pretest score on a depression questionnaire (i.e., the CES-D) indicated the potential presence of depression (the individual's score was 40, whereas the cutoff score in a representative adult German sample is 28; Hautzinger & Bailer, 1993). In the COGITO study, participants completed up to six daily sessions a week during a period of about 6 months. Among other variables, 11 depression-related symptoms were rated on these sessions, using either an 8-point scale or a 4-point scale (see Table 1) and with higher scores indicating stronger symptom presence. One hundred measurements

are available for the selected participant. Seventy-eight of the adjacent measurements were 1 day apart; 12 were 2 days apart; six were 3 days apart; and three were 4 or more days apart. There were no missing values. In Figure 1, the complete time series of each symptom is plotted.

Constructing a VAR (1)-based network model for these data, we can map the day-to-day symptom dynamics. This network representation can also reveal causal effects, provided one has included all relevant variables in the analysis. We will assume this condition is satisfied in the following, although ideally this should be checked thoroughly. In addition, the computation of centrality metrics allows identifying symptoms that seem to be crucial in increasing or maintaining the depression syndrome. Indeed, using the network figure, symptoms that are likely to be more important than others can be detected as well as particularly strong connections between symptoms. Ultimately, the analysis could thus provide a basis for possible interventions: Therapists can target core symptoms that incite many other symptoms when activated or aim to decouple two highly connected symptoms.

A network analysis based on VAR(1)

In this section, we elaborate on how to perform a VAR(1)-based network analysis, using the example data introduced previously. First, we briefly recapitulate VAR(1) analysis. Next, we discuss the network figure and the computation and interpretation of network characteristics. Finally, we cover the two problems introduced: comparability and unique effects.

The VAR(1) model

Given M stationary (i.e., the joint distribution is time invariant; Lütkepohl, 2005) time series with T equidistant measurement occasions, a VAR(1) model captures the lag-one relations among the M time series (Hamilton, 1994; Lütkepohl, 2005). The model separately regresses each variable at timepoint t on all variables at the previous timepoint t-1 (including the variable itself). An innovation term is added, referring to the part of the variable that cannot be predicted on the basis of the scores at the previous timepoint. Hence, the VAR(1) model can be written in vector notation as follows:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}\mathbf{y}_{t-1} + \mathbf{u}_t. \quad (1)$$

The Mx1 vectors $\mathbf{y}_t$ and $\mathbf{y}_{t-1}$ contain the values of the M variables at occasions t and t-1, respectively, and $\mathbf{c}$ is a Mx1 vector holding the intercepts. The MxM matrix $\mathbf{\Phi}$ contains the VAR(1) regression coefficients, with φ_{jk} indicating the regression weight of variable k when predicting variable j. On the diagonal of this matrix, the autoregressive (AR) parameters can be found, indicating how much each variable affects itself across time. The off-diagonal elements are the cross-regressive parameters, specifying the effect of one variable at time t-1 on another variable at time t. To ensure stationarity, the eigenvalues of $\mathbf{\Phi}$ (possibly complex numbers) should have a modulus smaller than 1 (Lütkepohl, 2005). The Mx1 vector $\mathbf{u}_t$ holds the innovations at time t. The innovations are assumed to be normally distributed with a zero mean vector and a variance covariance matrix $\mathbf{\Sigma}$, the inverse of which is sometimes referred to as the concentration matrix (Wild et al., 2010). The innovations cannot be serially correlated (i.e., across time), but instantaneous correlations (i.e., at the same timepoint) are allowed for. Hence, it can be concluded that the covariances of the variables depend on the regression coefficients as well as on the covariances of the innovations (see Lütkepohl, 2005).

Different VAR(1) estimation procedures are available. One can for example apply multivariate least squares (LS) estimation or formulate the model in state-space format and use associated estimation techniques (for details on these methods see, e.g., Hamilton, 1994; Lütkepohl, 2005). Moreover, techniques have been proposed to trim nonsignificant effects to obtain a sparse model. One way of arriving at a sparse model is using a stepwise procedure in which one starts from an empty model and sequentially adds edges that are significant. Alternatively, one may start from a full model and sequentially delete nonsignificant edges. A third strategy is to build several models of different complexities (amount of nonzero edges) and select the optimal one using an information criterion (Wild et al., 2010). Furthermore, lasso-based procedures exist that perform edge selection during model estimation (Abegaz & Wit, 2013; Rothman, Levina, & Zhu, 2010). Although an advantage of such sparse models or trimming procedures is that the obtained networks become less complicated, a disadvantage is that the obtained model may strongly depend on the procedure used (for examples in standard regression analysis, see Bondell & Reich, 2008; Chong & Jun, 2005). More fundamentally, however, the problems raised in this article still apply, irrespective of the specific trimming procedure used, as we will discuss later.

In this article, we made use of multivariate LS estimation, which yields identical results as a series of univariate ordinary least squares (OLS; Lütkepohl, 2005). It should be noted however that T has to be large enough to obtain good estimates (typically larger than 50, although the number of timepoints needed will probably be larger when the number of variables increases; see, e.g., Rosmalen, Wenting, Roest, de Jonge, & Bos, 2012; Wild et al., 2010). An interesting property of OLS is that it is an

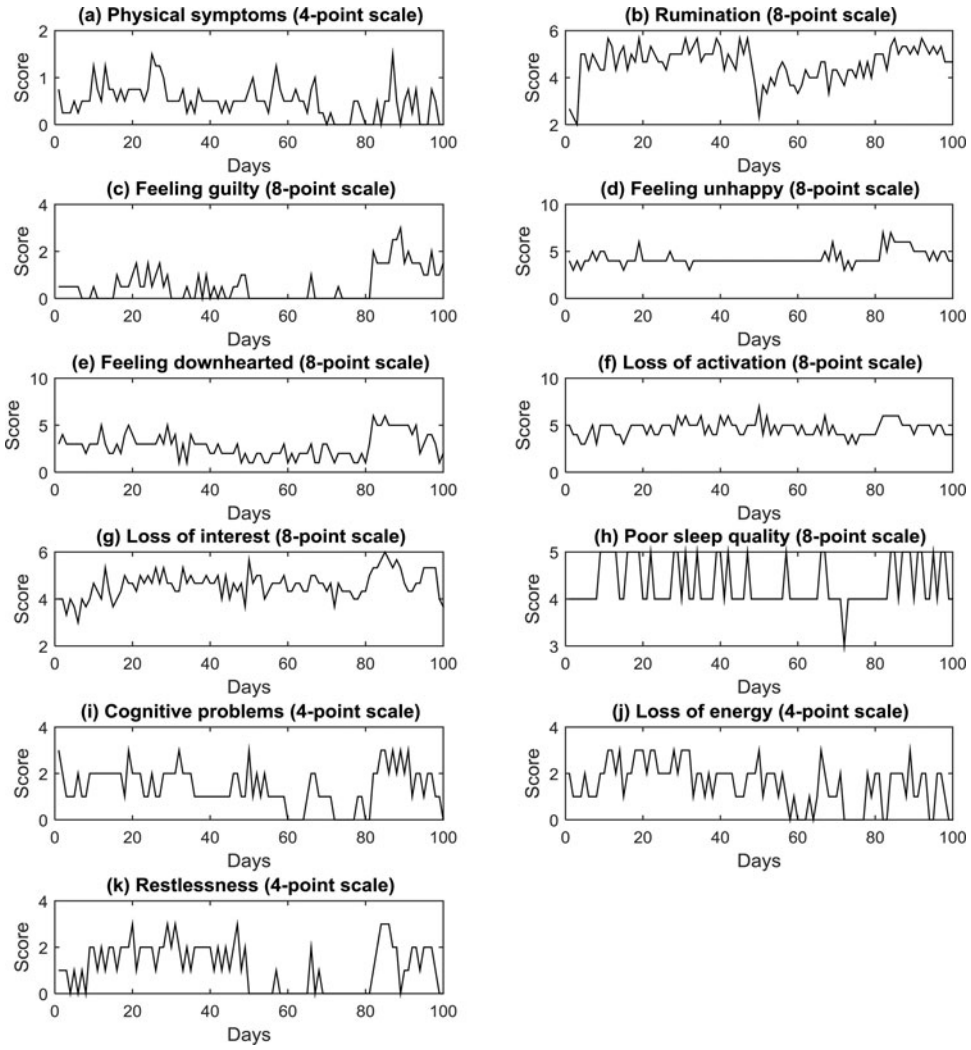


Figure 1. Time series plots of the 11 depression-related symptoms on 100 close-to-daily sessions.

affine equivariant estimator, which means that any affine transformation of the data (rotation, translation, rescaling, or any combination thereof) leads to a change in the parameter estimates but does not alter the percentage of variance explained or the obtained p values of the regression coefficients (Maronna, Martin, & Yohai, 2006).

Table 2 shows the resulting regression coefficients for our example data.[1] It can be concluded that the autoregressive coefficients are, in general, larger than the cross-regressive parameters. With regard to testing for Granger causality, only a few effects are significant (at a 5% level), which might be partly due to the relative small number of measurements given the number of variables. (Note that the number of significant coefficients further decreases after a Bonferroni correction for multiple comparisons.)

Network analysis

The network analysis is based on the regression coefficients in $\mathbf{\Phi}$, discarding the intercepts and innovations. To visualize the network (following the approach implemented in the R package qgraph; Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012), each variable is considered a node or vertex. The regression coefficients constitute the edges between the nodes. Specifically, each regression weight φ_{jk} is depicted as a directed link from node k pointing to node j. The edge thickness or strength reflects the size of the regression weight. Note that the qgraph package provides different options to clean up the graph (e.g., only show significant edges or edges that surpass a certain threshold), while the lasso-based technique

[1] We checked for the absence of serial dependencies in the residuals by means of lag length specification analysis and the Portmanteau test (Brandt & Williams, 2007; Lütkepohl, 2005). Lag length specification analysis was done by means of the Bayesian information criterion (BIC). The BIC value was minimal for a VAR model of order 1, indicating that a VAR(1) model is sufficient to capture the temporal dependencies (i.e., the BIC amounted to −9.40, −5.57, −1.74, 1.66, 4.10, 4.68, and 1.77 for models of order one up to seven). For the Portmanteau analysis, we used a lag of 7, as one might suspect the presence of a weekly cycle in the residuals of the VAR(1) model. The value of the adjusted test statistic, $\chi^2(726)=763.04$, and associated p value of .17 indicate that the null hypothesis of no serial correlation up to the tested lag cannot be rejected.

Table 2. The raw VAR(1) regression coefficients for the example data.

	Predictor										
Criterion	Physical symptoms	Rumination	Feeling guilty	Feeling unhappy	Feeling downhearted	Loss of activation	Loss of interest	Poor sleep quality	Cognitive problems	Loss of energy	Restlessness
Physical symptoms	.37**	−.02	.01	.01	.02	.02	−.04	−.11	−.07	.13*	.07
Rumination	−.17	.48**	.04	−.05	.04	−.10	.21	−.15	.09	−.02	.10
Feeling guilty	−.09	.13	.45**	.05	.10	−.05	.22*	−.31*	.06	−.07	.09
Feeling unhappy	−.28	.07	.27	.04	.12	.10	.32*	−.02	−.01	−.09	−.06
Feeling downhearted	−.46	.15	.31	.16	.33*	−.11	.18	−.21	.17	.09	.06
Loss of activation	−.02	.05	.16	.12	−.13	.09	.18	.17	−.05	.03	.22
Loss of interest	−.04	.08	.19	.04	−.08	.23*	.25*	−.12	−.03	−.06	.13
Poor sleep quality	.21	.12	.09	−.09	.04	−.06	.03	.05	.15	−.08	.04
Cognitive problems	.04	.03	.02	.16	.11	−.09	−.06	−.17	.25*	.14	.27*
Loss of energy	.32	.05	−.10	−.09	.03	−.05	−.10	.32	.04	.23	.29*
Restlessness	.03	.42	−.01	−.12	.14	.01	.17	−.30	.32*	−.01	.30*

*$p < .05$; **$p < .0004$ (Bonferroni correction).

is provided in the graphicalVAR package. The color of the edge indicates the sign of the regression weight, with green and red denoting positive and negative weights, respectively. We will refer to this network as the raw network because it is based on the raw VAR(1) regression coefficients.

Panel (a) of Figure 2 displays the raw network figure for the example data. Given that the variables are all symptoms of depression, one would intuitively expect mainly positive effects among them, with the activation of one symptom inciting the activation of others and vice versa. Although this hypothesis holds for some links (e.g., the effect of poor sleep quality on loss of energy and the effect of feeling guilty on feeling downhearted, to name the two most striking ones), many edges are red, indicating negative effects, which seems counterintuitive (e.g., the negative effect of physical symptoms on downhearted, which means that more physical symptoms come with feeling less downhearted). This observation will be discussed in more detail. For completeness, a network displaying the correlations between the innovations (computed on the basis of the regression residuals) can be found in panel (b) of Figure 2. According to Wild et al. (2010), this network can be called the partial contemporaneous correlations network.

Besides inspecting each single edge, the network can be further analyzed by calculating network characteristics (see Newman, 2010, for an introductory discussion). These characteristics are computed on the basis of the edges and thus on the regression weights. Examples are centrality measures (see, e.g., Bringmann et al., 2013), community structure characteristics, and interconnectedness measures (see, e.g., Pe et al, 2015). In this article, without loss of generality, we focus on centrality measures. These measures identify the most influential nodes of a network—that is, the nodes that account for an important part of the network flow. For our example data, centrality measures are thus useful to identify the core symptoms maintaining the depressed state. Since the importance of a node can be defined in many ways, a number of centrality metrics were developed. Here, we will discuss four of them: instrength, outstrength, betweenness, and closeness (cf. Opsahl, Agneessens, & Skovetz, 2010). They can be easily calculated in the R package qgraph.

Instrength and outstrength measures define node importance in terms of the strength of the direct relations of a node to all other nodes. The instrength of a particular node refers to the sum of the absolute values of the edge weights directed at this node. Outstrength is computed as the sum of the absolute values of all edge weights leaving from this node. In other words, a symptom with a high instrength is directly activated by many other symptoms, and a symptom with a high outstrength tends to directly activate many other symptoms.

Betweenness and closeness are based on the notion of the shortest path between two nodes. A path may pertain to the direct link between them but also to routes that include intermediary nodes. The length of a particular path is computed as the sum of the absolute value of the inverse weights of the edges that constitute the path. The shortest path out of all possible paths between two nodes is the one that minimizes this sum. The betweenness centrality of a node equals the number of shortest paths passing through the node (Opsahl et al., 2010). Nodes with a relatively high betweenness index can be seen as "gatekeepers" because they are situated on the paths connecting the less central nodes (Kolaczyk, 2009, Chapter 4). In turn, closeness centrality is defined as the inverse of the sum of the shortest path lengths of a node to all other nodes (Opsahl et al., 2010). A high closeness centrality indicates that this particular node has a larger influence

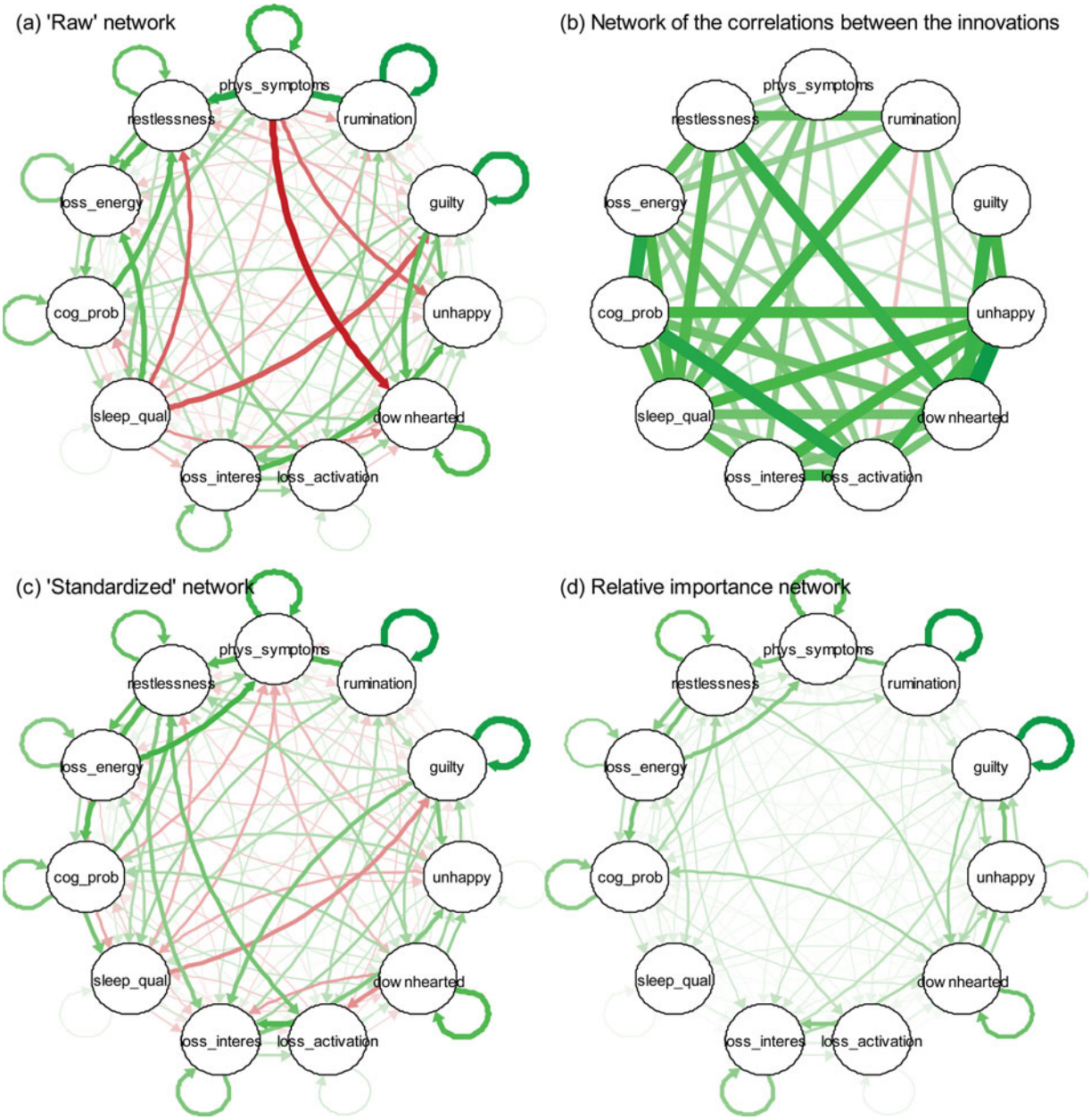


Figure 2. Network figures of the example data. The following symptom labels are used: phys_symptoms = physical symptoms; rumination = rumination; guilty = feeling guilty; unhappy = feeling unhappy; downhearted = feeling downhearted; loss_activation = loss of activation; loss_interest = loss of interest; sleep_quality = poor sleep quality; cog_prob = cognitive problems; loss_energy = loss of energy; restlessness = restlessness. Panel (a) shows the raw network; panel (b) shows a network figure of the correlations between the innovations; panel (c) displays the standardized network. The network in panel (d) is based on the LMG metric, which is a measure of relative importance.

on the other nodes of the network than does a node with a lower closeness centrality, in that, for example, more nodes are influenced or the same nodes are more strongly affected. Closeness differs from outstrength in that closeness takes into account all paths that leave from a node, rather than only the direct ones.

Two remarks regarding these centrality measures are in order. First, a centrality value of a node should be interpreted in a relative way, by comparing it to the values of the other nodes of the same network. Second, the different measures may point at different nodes to be most influential because they rely on different definitions of importance (Borgatti, 2005; Kolaczyk, 2009).

Figure 3 displays the values of the four centrality measures for the 11 depression-related symptoms in our example data. Whereas downhearted and restlessness have the highest instrength (i.e., many other symptoms activate feeling restless and downhearted), physical symptoms and sleep quality have the highest outstrength (i.e., physical symptoms and sleep quality activate many other

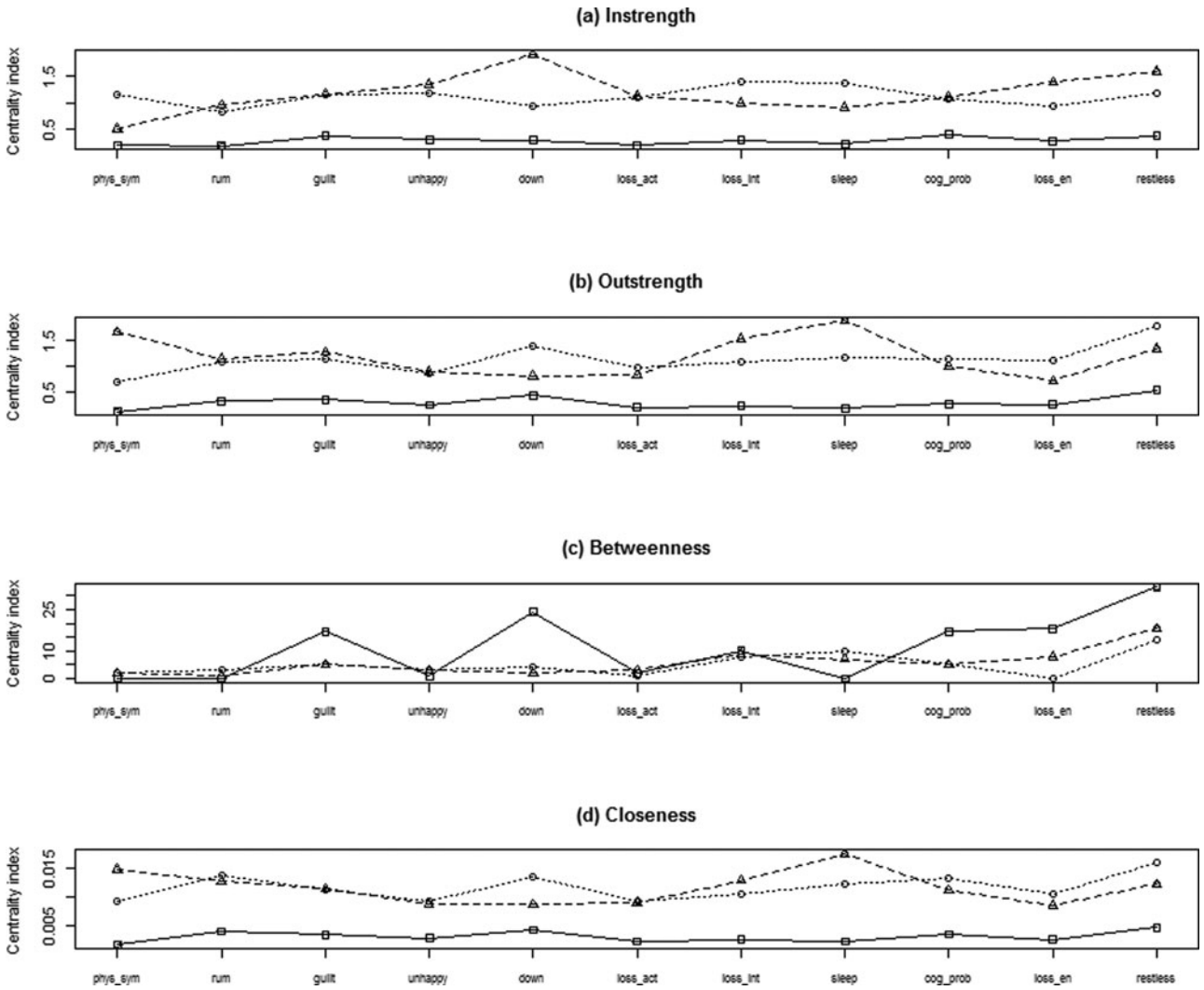


Figure 3. Four centrality measures for the 11 depression-related symptoms. The following labels are used: phys_sym = physical symptoms; rum = rumination; guilt = feeling guilty; unhappy = feeling unhappy; down = feeling downhearted; loss_act = loss of activation; loss_int = loss of interest; sleep = poor sleep quality; cog_prob = cognitive problems; loss_en = loss of energy; restless = restlessness. The measures are based on the raw (dashed lines), standardized (dotted lines), and relative importance–based (solid lines) network figures.

symptoms in the network). Using the closeness measure yields results similar to the outstrength index (i.e., physical symptoms and sleep quality influence more symptoms than other nodes, or they influence other nodes more strongly), which could be expected given the definitions above. The betweenness metric selects restlessness as the most influential symptom in the sense that it connects other symptoms more than other nodes do.

Two problems: Comparability and unique effects

In the previous section, we elucidated that the edges are crucial when interpreting the network figure as well as when computing network characteristics. In this article, we argue that one should be cautious, however, when using the raw VAR(1) coefficients because of two problems that are well known in the context of standard regression analysis.

First, the edges might not be comparable in terms of relation strength due to scale or variance differences between the variables. Indeed, the variances of the variables that serve as predictor and criterion may have a large effect on the size of the associated regression coefficients. For example, a predictor that is expressed as a proportion will have a variance smaller than 1. If this predictor is used to predict a criterion measured on a scale from 0 to 100, one expects to obtain a large regression coefficient if both variables are at least somewhat related. If the same criterion is regressed on another predictor that is equally related to the criterion and measured on the same scale as the criterion, the obtained regression coefficient will automatically be smaller, although the link is equally strong. This example demonstrates that differences in the

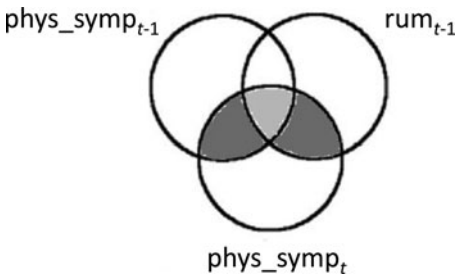


Figure 4. A visual representation of the difference between the unique direct effects of predictors (the dark grey areas) and their shared effect (the light grey area). For purposes of illustration, two depression-related symptoms were used: phys_symp = physical symptoms; rum = rumination. The predictors are measured at the previous timepoint *t*-1, and the criterion variable at timepoint *t*.

edges might be due to scale differences only and thus that the edges with the largest width do not necessarily reflect the strongest Granger causal processes. Hence, network characteristics that are computed on the basis of these raw regression coefficients will be misleading as well. This problem not only occurs when variables are measured with different scales; it can also affect results when variables are measured on the same scale but differ considerably in variance (see, e.g., the variances of the symptoms feeling downhearted and sleep quality in Table 1).

Second, it is important to realize that regression coefficients capture the unique direct effects of the predictors in that they reflect which part of the variance of the criterion is uniquely predicted by a specific variable. The predictive variance shared by variables—the part of the criterion variance that can be explained by multiple predictors—is thus not taken into account. This phenomenon is clarified in Figure 4. The circles represent the variances of the predictors (phys_symp$_{t-1}$ and rum$_{t-1}$) and the criterion (phys_symp$_t$). The intersection of a single predictor and the criterion, which is colored dark grey in Figure 4, represents the explained variance attributable to the unique direct effect of that predictor, whereas the intersection of all three variables, colored light grey, indicates the shared effect of both predictors (Cohen, Cohen, West, & Aiken, 2003, Chapter 3). Thus, the more the predictors are correlated (i.e., the intersection between the predictors in Figure 4 becomes larger), the larger is the effect that is ignored by only taking the unique direct effects into account. However, the larger these shared influences, the more important it is to incorporate them in the analysis as they may give more insight into the temporal relations among the variables and qualify the Granger causal conclusions that are based on the unique direct effects only. Moreover, exclusively focusing on unique effects may also explain the presence of counterintuitive negative links in networks, where in fact the total effect of the predictor on the criterion is positive (such as we obtained for our example data). Hence, we conclude that the importance of a predictor for understanding a criterion may be underestimated in VAR(1) as it was previously used in the literature. This problem is sometimes referred to as overcontrol, in that the shared effects are removed from the estimated effect of a particular predictor (Cohen et al., 2003).

VAR estimation methods that trim some of the edges sidestep to some extent the shared effects issue as substantial amounts of shared variance between a number of variables will be less likely to occur. However, the trimming introduces a different problem: In case of highly correlated variables, which edges are trimmed is often arbitrary and depends on sampling fluctuations (Bondell & Reich, 2008). This is unsatisfactory if we take the perspective of an applied researcher or therapist. To illustrate this, consider the example of two variables (feeling guilty and rumination) that jointly explain a large proportion of variance of a third variable (negative affect). Using an edge-trimming method, one or both variables will probably not enter the model because the additional contribution on top of the other variable will be rather small. Yet both variables might be interesting targets for an intervention, and the one not selected by the heuristic procedure might even be a more convenient target (e.g., feeling guilty).

If we reexamine the raw network in Figure 2, it becomes clear that both these problems might have hindered and even distorted our interpretation. Regarding the comparability issue, from the figure, one is inclined to think that the negative edge from physical symptoms to downhearted is an important link in the obtained network. Yet based on the scale and variance differences of the variables in Table 1, we could already have expected this edge to have the largest width. Indeed, physical symptoms, measured on a 4-point scale, is the variable with the smallest variance, while downhearted, measured on an 8-point scale, has the largest variance, leading to a large raw regression weight. Similarly, the small variance of physical symptoms also explains the high outstrength and closeness values of this variable (see Figure 3) and sheds doubt on the conclusion that this symptom is very influential.

Furthermore, to illustrate that the example data might contain many shared effects that are ignored in the network, Figure 5 displays the R^2, the percentage of explained variance, for each criterion. We also calculated the sum of the squared semipartial correlations (i.e., the percentage of uniquely explained variance) of the predictors for each criterion, which gives an indication to what extent the R^2 values are attributable to the depicted unique direct (i.e., potential Granger causal) effects. Figure 5 clearly shows that the sum of these unique effect shares do not add up to the R^2. However, in the network figure and the associated centrality measures, only these unique effects are taken into account, implying that a lot of information is ignored.

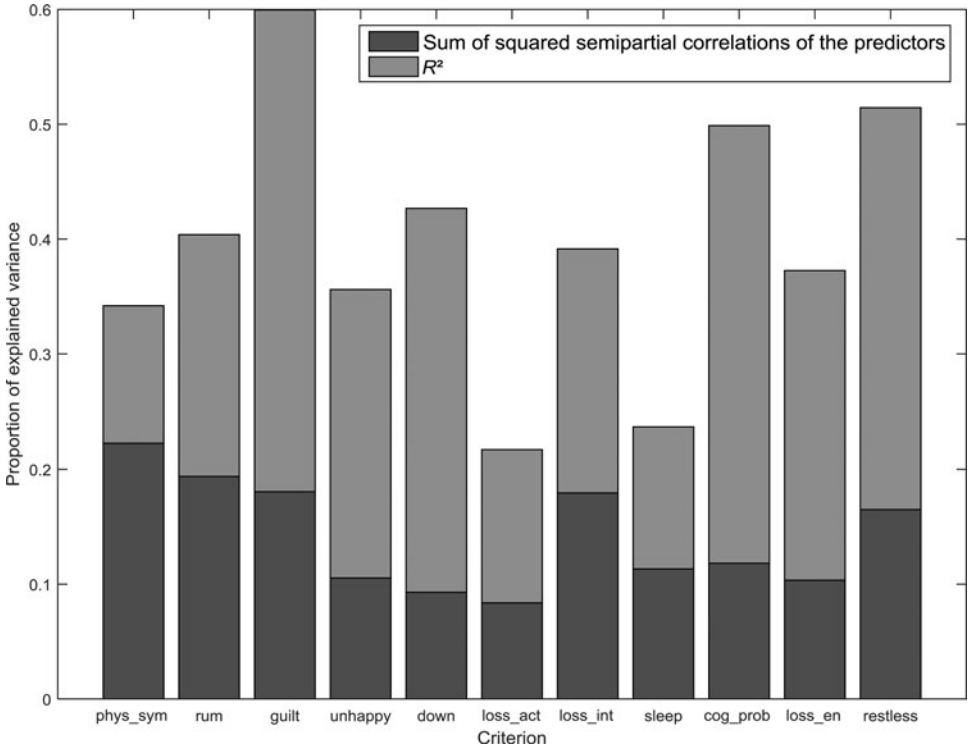


Figure 5. Bar plot of the proportion of explained variance R^2 for the 11 depression-related symptoms as criterion variables. The dark grey bars display the sum of the squared semipartial correlations of all predictors. Note that the interpretation of these sums can sometimes be intricate (for more details, see Cohen et al., 2003). The 11 symptoms: phys_sym = physical symptoms; rum = rumination; guilt = feeling guilty; unhappy = feeling unhappy; down = feeling downhearted; loss_act = loss of activation; loss_int = loss of interest; sleep = poor sleep quality; cog_prob = cognitive problems; loss_en = loss of energy; restless = restlessness.

A solution to the comparability problem: A network based on standardized VAR(1) weights

To ensure the comparability of the VAR(1) regression coefficients and thus of the edges, one can use the solution that is provided in classical regression analysis: Standardize the coefficients. Indeed, standardization takes care of differences in variances between variables, implying that one can directly compare the regression weights to derive which of the predictors has the largest unique direct effect and thus possibly the strongest Granger causal influence. The standardized weights $\mathbf{\Phi}^*$ can be obtained by applying the regression analysis to the z scores for each variable. Alternatively, one can simply reparametrize the raw regression coefficients:

$$\mathbf{\Phi}^* = \begin{pmatrix} \varphi_{11}\frac{s_1}{s_1} & \cdots & \varphi_{1M}\frac{s_M}{s_1} \\ \vdots & \ddots & \vdots \\ \varphi_{M1}\frac{s_1}{s_M} & \cdots & \varphi_{MM}\frac{s_M}{s_M} \end{pmatrix}, \qquad (2)$$

where s_i denotes the standard deviation of variable i.

Although the affine equivariance property of the OLS estimator ensures that the different R^2 values are not affected by the standardization, it is important to keep in mind that this transformation does call for a new interpretation of the VAR(1) regression coefficients. Luskin (1991) nicely elaborates this new interpretation for the case of classical multiple regression: "The standardized coefficient measures the proportion of the greatest likely variation in y that can be accounted for by the greatest likely variation in x" (p. 1035), holding the other variables constant.

To illustrate the effect of using the standardized rather than the raw VAR(1) coefficients, we return to our data example. Panel (a) of Figure 6 shows a scatter plot of the raw versus the standardized regression coefficients for these data. Although both sets of coefficients are very strongly correlated ($r = .91$), one can discern some clear differences, especially for the regression coefficients with the highest absolute values. Some of these values are much lower after standardization, which implies that the size of the coefficient was indeed due to variance differences among the variables. Because the coefficients with the largest absolute values are the ones that are the most striking in the network figure and have a huge effect on the obtained centrality measures, it is no surprise that the conclusions based on the standardized network (Figure 2, panel (c)) and the associated centrality measures (Figure 3, dotted line) differ from those derived from the raw network. Regarding the network links, two

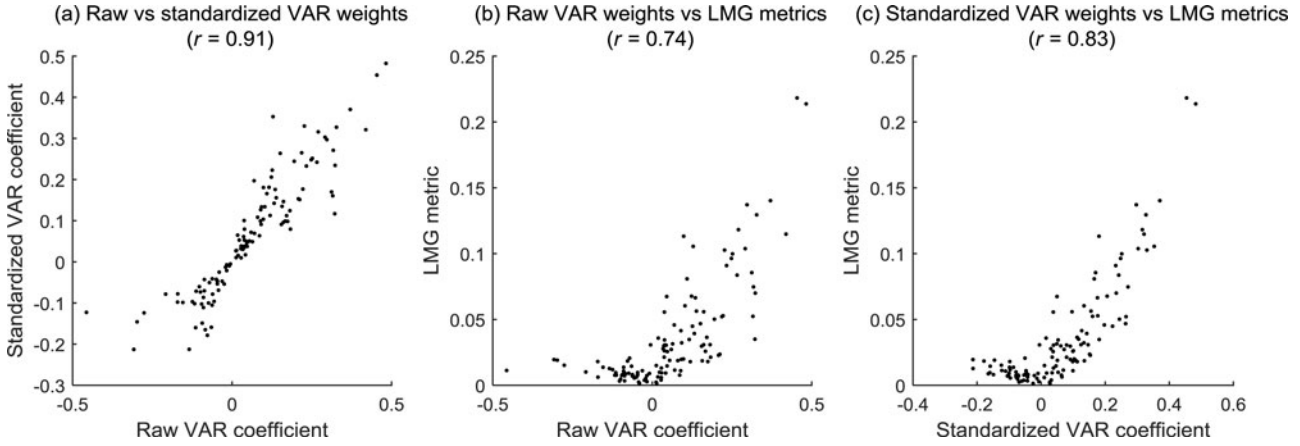


Figure 6. Scatter plots of the raw lag-one vector autoregressive (VAR(1)) coefficients, standardized VAR(1) coefficients, and LMG measures (Lindeman, Merenda & Gold, 1980) for our example data.

of the counterintuitive negative effects fade in the standardized network compared to the raw network (i.e., the negative effect of physical symptoms on downhearted and that of sleep quality on restlessness). Regarding centrality, instrength now points at the importance of loss of interest and sleep quality. The outstrength and closeness measures select restlessness and feeling downhearted as the core symptoms. Only for betweenness centrality, similar conclusions are obtained, with restlessness being the most influential symptom in the network.

A solution to the unique effects problem: A network based on relative importance metrics

It is well known in classical multiple regression analysis that both the raw and the standardized regression coefficients reflect only the unique direct effects of the predictors, ignoring their shared effects. At first sight, an easy solution to this problem would be to inspect the correlations among each of the predictors and the criterion as they reflect the total effect of a predictor and thus also the shared effects. However, this does not really help as it does not reveal which part of the total effect is shared. To explain why this is important, we again use the example of a therapeutic intervention in which one wants to treat those symptoms that cause other symptoms across time. Imagine that two predictors both strongly correlate with the criterion. These results can imply that they both have a strong direct unique effect, which makes it meaningful to target both variables in the therapeutic intervention. However, another possibility is that both predictors are substantially correlated and thus that an important part of their effect on the criterion is shared. In such cases, an intervention on one of the variables may be enough to reverse the effect. Running additional analyses is necessary to discriminate between both possibilities.

This issue resulted in a strand of literature regarding relative importance metrics, which aim to reflect the shared effects in a better way. A range of relative importance definitions and associated exploratory measures were proposed, each having its advantages and disadvantages (for reviews and comparisons, see Budescu, 1993; Johnson, 2000; Johnson & LeBreton, 2004). In line with Grömping (2006), we will adhere to the relative importance definition of Johnson and LeBreton (2004) and use a metric that estimates the proportionate contribution each predictor makes to R^2. This contribution equals the sum of the unique direct effect and a share of the combined effect of the predictors. Other relative importance metrics were proposed (e.g., zero-order correlation, product measure, and t statistic; see Johnson & LeBreton, 2004) but will not be considered here since they do not add up to R^2 and thus do not comply with this definition. Three options remain: the LMG metric (as Grömping, 2007, named it after its developers Lindeman, Merenda & Gold, 1980, which boils down to an average squared semipartial correlation), the general dominance measure (Azen & Budescu, 2003), and relative weights (Johnson, 2000). The first two are computationally equivalent (Azen & Budescu, 2003). We mainly opt for the LMG metric because it is well investigated (e.g., for information on the sampling distribution, see Grömping, 2007) and because the R package relaimpo is available for the calculations (Grömping, 2006).

The LMG score gives an estimate of the gain in R^2 due to a particular predictor by averaging the increase in R^2 due to this predictor over all possible orderings of predictors. In case of M predictors, there are $M!$ possible orderings. Some orderings will result in the same semipartial correlations. In particular, the semipartial correlation does not change when the ordering of all variables entered prior to or after the variable of interest changes. Hence, the LMG metric for predictor y_{1t-1} and criterion

y_{kt} is computed in the following way:

$$LMG\left(y_{kt}y_{1t-1}\right) = \frac{1}{M!}\left(\begin{array}{l}(M-1)!r^2_{y_{kt}y_{1t-1}} + (M-2)!r^2_{y_{kt}(y_{1t-1}.y_{2t-1})} + \ldots + (M-2)!r^2_{y_{kt}(y_{1t-1}.y_{Mt-1})} \\ +2!(M-2-1)!r^2_{y_{kt}(y_{1t-1}.y_{2t-1}y_{3t-1})} + \ldots + 2!(M-2-1)!r^2_{y_{kt}(y_{1t-1}.y_{2t-1}y_{Mt-1})} \\ +2!(M-2-1)!r^2_{y_{kt}(y_{1t-1}.y_{3t-1}y_{4t-1})} + \ldots + 2!(M-2-1)!r^2_{y_{kt}(y_{1t-1}.y_{3t-1}y_{Mt-1})} \\ +\ldots + (M-1)!r^2_{y_{kt}(y_{1t-1}.y_{2t-1}y_{3t-1}\ldots y_{Mt-1})}\end{array}\right)$$
$$= \frac{1}{M!}\sum_{S\subseteq\{2,\ldots,M\}} s!(M-s-1)!r^2_{y_{kt}(y_{1t-1}.S)}, \quad (3)$$

with $r^2_{y_{kt}(y_{1t-1}.y_{jt-1})}$ indicating the squared semipartial correlation between the criterion variable y_{kt} and the predictor y_{1t-1} from which y_{jt-1} has been partialed, S representing a subset of the predictor variables (except the one of interest) entered prior to y_{1t-1}, and s denoting the number of variables of the subset (see Grömping, 2007, for further details). The LMG metric nicely decomposes the R^2 because the sum of the LMG values of all predictors equals the R^2. The LMG metric thus results in positive values for each predictor as it does not take the direction of the different unique and shared effects into account. The calculation is performed separately for each criterion. Note that scale or variance differences among the variables do not play a role since the LMG metric is based on semipartial correlations.

Hence, to solve the unique effects problem, we propose to construct the network on the basis of the relative importance metrics rather than the VAR(1) regression coefficients (for an illustration of this idea in a cross-sectional study and thus without using VAR, see Robinaugh, LeBlanc, Vuletich, & McNally, 2014). In such a relative importance–based network, the edges give an indication of the amount of variance of the receiving node that can be attributed to the sending node, while taking into account shared effects. Since LMG values are by definition positive, all edges will be colored green. Moreover, the LMG values vary less in size than the raw and standardized regression weights because their sum per receiving node amounts to 1 at maximum. Therefore, the thickness of the edges will vary less as well as the centrality values.

Turning to our example data, the scatter plots in Figure 6 show clear differences between the LMG edges and the ones that constitute the raw and standardized network. The shapes of these scatter plots reveal that all edges that were negative in the raw and standardized networks have very low LMG values (e.g., the effect of poor sleep quality on feeling guilty), indicating that these effects are of low predictive importance. The large regression coefficients of these variables can therefore be interpreted as suppression effects. Regarding the positive raw and standardized coefficients, the standardized coefficients seem to be more in line with the LMG metric as the values are less dispersed in panel (c) compared to panel (b) of Figure 6. Examining the relative importance–based network figure (the last panel of Figure 2), some important links can be discerned—for example, the link pointing from restlessness to loss of activation. In contrast to the raw and standardized values, the relative importance–based centrality values (Figure 3, solid line) suggest that the somatic component of depression (sleep problems and physical symptoms) are not influential in this network. Instead, the most important symptoms seem to be negative feelings (downhearted and guilty) and restlessness.

Discussion

The dynamic network approach is a useful tool to gain insight into causal relations among variables at an intraindividual level (if warranted by the design of the study). However, we argued in this article that one should be cautious when the networks are built on raw VAR(1) regression coefficients, as is often done. Specifically, we pointed out two important problems. The first problem pertains to the comparability of the edges, where we conjectured that raw VAR(1) regression coefficients and thus also the size of the Granger causal effects are not comparable in case of scale or variance differences. To deal with this problem, we recommended using the standardized VAR(1) regression weights instead. The second problem is that both the raw and standardized VAR(1) regression weights reflect only the unique direct effects of the variables. To capture the shared effects of the variables as well, we recommend using the LMG metric, which is an exploratory approach to distribute the total percentage of explained variance of a criterion across the predictors. Analyzing empirical data of an individual who reported 11 depression-related symptoms across 100 days, we showed that it indeed matters which measure constitutes the edges of the network, in that different links pop out as the most striking ones in the network figures and in that the centrality measures do not pick the same most influential symptoms.

In the following paragraphs, we elaborate on three topics. First, how should one choose between the different metrics when analyzing data? Second, what are interesting avenues for future research? Third, what about other

regression-based methods for studying causal relations across time? Do they suffer from the same problems?

What to use: Standardized coefficients and/or LMG scores?

When deciding whether to draw and analyze a standardized network and/or a relative importance one, two questions can be helpful. First, one should determine whether information regarding the direction of the effects (positive or negative) is needed. Indeed, whereas the standardized network reveals whether a particular node up or down regulates other nodes, the LMG metric does not take the signs of the effects into account. Second, one should decide whether one is interested in the unique direct effects and thus in Granger causal processes or in the shared effects as well. The answer depends on the research question at hand, which thus should be formulated precisely. For example, when studying the emotional interplay between mother, father, and child during a conflict situation, unique direct effects will show which parent has the largest unique influence on the situation, whereas taking shared effects into account as well allows to get an idea of whether mother and father unite force. Similarly, when studying the phenomenology of depression in an individual, unique effects will show whether, for example, feelings of guilt or rumination have the largest unique influence on negative feelings. Instead, taking shared effects into account allows getting an idea whether feelings of guilt and rumination have a shared effect. If the latter is the case, it seems relevant from a practitioner's perspective to determine how the two symptoms are intertwined.

However, we also argued that more insights into the different effects of the variables and the underlying mechanisms can be obtained by inspecting the scatter plots of the raw, standardized, and relative importance–based edges (see Figure 6). On the basis of these plots, we concluded that the counterintuitive negative links in the raw network of the depression symptoms are likely due to large variance differences among the symptoms on the one hand and suppression effects on the other hand. Together, we conclude that rather than picking one network, it often will be more informative to report the three networks and draw conclusions that are justified from a specific perspective (e.g., a focal treatment of symptoms with strong Granger causal effects or a treatment that aims at understanding a net of intertwined variables, with the long-term goal to disentangle a vicious cycle). Inspecting multiple network figures may not be a satisfying solution to some; further research can help establish which network provides the most useful basis for intervention.

Avenues for future research

The proposed alternative edge measures may not be the ultimate ones, in that both the standardized coefficients and the LMG metric come with disadvantages as well. Regarding the use of standardized regression weights, we focused on *z*-score-based standardization. However, other standardization options have been proposed in the literature. For example, Gelman (2008) argues to scale by dividing by 2 times the standard deviation. Bring (1994) advocates using the partial or conditional standard deviation, which gives an indication of the dispersion of a predictor holding all other predictors constant. It can be estimated by regressing a particular predictor on the remaining predictors. Further research on the properties of these different standardization options seems necessary.

Regarding the LMG metric, the resulting scores are essentially univariate in that they are computed for each criterion separately (Johnson & LeBreton, 2004). From a network perspective, a multivariate version of importance metrics is called for that takes all equations at once into consideration as the different variables will always be correlated. Such multivariate relative importance measures would reflect the relative importance of a predictor for understanding changes in the full dynamic system. The work of LeBreton and Tonidandel (2008) may be a useful starting point for this endeavor.

A general limitation of the present article is the neglect of estimation uncertainty. Although estimates of the standard errors of all three edge measures can be computed, this estimation uncertainty is not taken into account when computing the centrality measures. This is an important direction for future research.

Finally, the issues discussed in the current article also apply to extensions of the VAR model. For example, we conjecture that one should also be cautious when interpreting the regression weights that are obtained using a multilevel regression model to time series data of multiple persons (Bringmann et al., 2013). Translating the proposed solutions to a multilevel context is not straightforward (e.g., different standardization options exist), however, and more research on the topic is needed.

Other regression-based methods for constructing dynamic networks

As stated in the introduction, alternative regression-based approaches for capturing dynamic relations exist. Specifically, one may use structural VAR models (SVAR; Lütkepohl, 2005, Chapter 9) or unified structural equation models (uSEM; Kim, Zhu, Chang, Bentler, & Ernst, 2007)

as these SVAR models are labeled within the SEM framework. The main difference between the reduced form VAR that we focused on and SVAR is the way they deal with contemporaneous relations between the variables, on top of the lagged ones. These contemporaneous relations are likely present because an external event in many cases affects multiple variables simultaneously or because measurement frequency is too low (i.e., if the gaps between the different timepoints are too high, quickly evolving lagged effects turn up as contemporaneous ones). To obtain unbiased parameters of the causal relations, one ideally models the contemporaneous relations as well (Gates, Molenaar, Hillary, Ram, & Rovine, 2010) by restricting the variance–covariance matrix of the innovations to be diagonal and including an additional term in Equation (1), expressing how well each variable can be predicted on the basis of the other variables at the same timepoint. This is not trivial, however, since assumptions about the presence or absence of some effects are needed to identify the model (Brandt & Williams, 2007; Lütkepohl, 2005); that is, not all instantaneous and lagged relations can be estimated simultaneously. However, more important regarding the topic of this article, including additional model parameters does not offer by itself a solution to the comparability and unique effects problems.

Conclusion

To conclude, we discussed two important problems when building an intraindividual network on raw VAR(1) coefficients, and we proposed two solutions. First, raw VAR(1) coefficients lack comparability, which can be solved by standardizing the coefficients. Second, whereas the regression weights only reflect the unique direct effects of the predictors, LMG scores allow exploring shared effects. It seems to us that in many empirical applications, it will be more informative to report the three networks and draw combined and informed conclusions rather than relying on one network alone.

Article information

Conflict of Interest Disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical Principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: Kirsten Bulteel is a doctoral research fellow with the Research Foundation–Flanders. This work was further supported by Grant GOA/15/003 from the Research Fund of KU Leuven and by IAP/P7/06 from the Interuniversity Attraction Poles program financed by the Belgian government. The COGITO study was funded by a grant of the Innovation Fund of the Max Planck Society to Ulman Lindenberger.

Role of the Funders/Sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments: The authors thank three anonymous reviewers and the associate editor for their comments on prior versions of this manuscript. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institutions is not intended and should not be inferred.

References

Abegaz, F., & Wit, E. (2013). Sparse time series chain graphical models for reconstructing genetic networks. *Biostatistics*, *14*, 586–599. doi:10.1093/biostatistics/kxt005

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, *8*, 129–148. doi:10.1037/1082-989X.8.2.129

Bondell, H. D., & Reich, B. J. (2008). Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. *Biometrics*, *64*, 115–123. doi:10.1111/j.1541-0420.2007.00843.x

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*, 55–71. doi:10.1016/j.socnet.2004.11.008

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

Bos, E. H., Hoenders, R., & de Jonge, P. (2012). Wind direction and mental health: A time-series analysis of weather influences in a patient with anxiety disorder. *BMJ Case Reports*. doi:10.1136/bcr-2012-006300

Brandt, P. T., & Williams, J. T. (2007). *Multiple time series models*. Thousand Oaks, CA: Sage Publications.

Bring, J. (1994). How to standardize regression coefficients. *The American Statistician*, *48*, 209–213. doi:10.1080/00031305.1994.10476059

Bringmann, L. F., Lemmens, L. H. J. M., Huibers, M. J. H., Borsboom, D., & Tuerlinckx, F. (2015). Revealing the dynamic network structure of the Beck Depression Inventory-II. *Psychological Medicine*, *45*, 747–757. doi:10.1017/S0033291714001809

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., … Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS ONE*, *8*, 1–13. doi:10.1371/journal.pone.0060188

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, *114*, 542–551. doi:10.1037/0033-2909.114.3.542

Chong, I.-G., & Jun, C.-H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, *78*, 103–112. doi:10.1016/j.chemolab.2004.12.011

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3th ed.). New York, NY: Routledge.

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*, 1–18. doi:10.18637/jss.v048.i04

Gates, K. M., Molenaar, P. C. M., Hillary, F. G., Ram, N., & Rovine, M. J. (2010). Automatic search for fMRI connectivity mapping: An alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM. *NeuroImage*, *50*, 1118–1125. doi:10.1016/j.neuroimage.2009.12.117

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, *27*, 2865–2873. doi:10.1002/sim.3107

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*, 424–438. doi:10.2307/1912791

Grömping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software*, *17*, 1–27. doi:10.18637/jss.v017.i01

Grömping, U. (2007). Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician*, *61*, 139–147. doi:10.1198/000313007×188252

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Hautzinger, M., & Bailer, M. (1993). *Allgemeine Depressions Skala* [General Depression Scale]. Weinheim, Germany: Beltz.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*, 1–19. doi:10.1207/S15327906MBR3501_1

Johnson, J. W., & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, *7*, 238–257. doi:10.1177/1094428104266510

Kim, J., Zhu, W., Chang, L., Bentler, P. M., & Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. *Human Brain Mapping*, *28*, 85–93. doi:10.1002/hbm.20259

Kolaczyk, E. D. (2009). *Statistical analysis of network data: Methods and models*. New York, NY: Springer.

LeBreton, J. M., & Tonidandel, S. (2008). Multivariate relative importance: Extending relative weight analysis to multivariate criterion spaces. *Journal of Applied Psychology*, *93*, 329–345. doi:10.1037/0021-9010.93.2.329

Lindeman, R. H., Merenda, P. F., & Gold, R. Z. (1980). *Introduction to bivariate and multivariate analysis*. Glenview, IL: Scott Foresman.

Luskin, R. C. (1991). Abusus non tollit usum: Standardized coefficients, correlations, and R^2s. *American Journal of Political Science*, *35*, 1032–1046. doi:10.2307/2111505

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin, Germany: Springer.

Maronna, R. A., Martin, R. D., Yohai, V. J. (2006). *Robust statistics : Theory and methods*. Chichester, UK: John Wiley & Sons Ltd.

Newman, M. E. J. (2010). *Networks: An introduction*. New York, NY: Oxford University Press.

Opsahl, T., Agneessens, F., & Skovetz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*, 245–251. doi:10.1016/j.socnet.2010.03.006

Pe, M. L., Kircanski, K., Thompson, R. J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., … Gotlib, I. H. (2015). Emotion-network density in major depressive disorder. *Clinical Psychological Science*, *3*, 292–300. doi:10.1177/2167702614540645

Pe, M. L., & Kuppens, P. (2012). The dynamic interplay between emotions in daily life: Augmentation, blunting, and the role of appraisal overlap. *Emotion*, *12*, 1320–1328. doi:10.1037/a0028262

Robinaugh, D. J., LeBlanc, N. J., Vuletich, H. A., & McNally, R. J. (2014). Network analysis of persistent complex bereavement disorder in conjugally bereaved adults. *Journal of Abnormal Psychology*, *123*, 510–522. doi:10.1037/abn0000002

Rosmalen, J. G. M., Wenting, A. M. G., Roest, A. M., de Jonge, P., & Bos, E. H. (2012). Revealing causal heterogeneity using time series analysis of ambulatory assessments: Application to the association between depression and physical activity after myocardial infarction. *Psychosomatic Medicine*, *74*, 377–386. doi:10.1097/PSY.0b013e3182545d47

Rothman, A. J., Levina, E., & Zhu, J. (2010). Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics*, *19*, 947–962. doi:10.1198/jcgs.2010.09188

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2009). On the relation of mean reaction time and intraindividual reaction time variability. *Psychology and Aging*, *24*, 841–857. doi:10.1037/a0017799

Schmitz, B., & Skinner, E. (1993). Perceived control, effort, and academic performance: Interindividual, intrainindividual, and multivariate time-series analysis. *Journal of Personality and Social Psychology*, *64*, 1010–1028. doi:10.1037/0022-3514.64.6.1010

Snippe, E., Bos, E. H., van der Ploeg, K. M., Sanderman, R., Fleer, J., & Schroevers, M. J. (2015). Time-series analysis of daily changes in mindfulness, repetitive thinking, and depressive symptoms during mindfulness-based treatment. *Mindfulness*, *6*, 1053–1062. doi:10.1007/s12671-014-0354-7

van Gils, A., Burton, C., Bos, E. H., Janssens, K. A. M., Schoevers, R. A., & Rosmalen, J. G. M. (2014). Individual variation in temporal relationships between stress and functional

somatic symptoms. *Journal of Psychosomatic Research*, *77*, 34–39. doi:10.1016/j.jpsychores.2014.04.006

Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, *44*, 1349–1360. doi:10.1017/S0033291713001979

Wild, B., Eichler, M., Friederich, H.-C., Hartmann, M., Zipfel, S., & Herzog, W. (2010). A graphical vector autoregressive modelling approach to the analysis of electronic diary data. *BMC Medical Research Methodology*, *10*, 28. doi:10.1186/1471-2288-10-28