Measurement targets for network constructs in psychopathology

van Bork, R.; Lunansky, G.; Borsboom, D.

DOI
10.1016/j.measurement.2024.114643
Publication date
2024
Document Version
Final published version
Published in
Measurement

Measurement 232 (2024) 114643

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Measurement targets for network constructs in psychopathology

Riet van Bork [a,b,*], Gabriela Lunansky [c], Denny Borsboom [a]

[a] University of Amsterdam, The Netherlands
[b] Center for Philosophy of Science, University of Pittsburgh, United States
[c] Amsterdam Public Health Institute, Amsterdam University Medical Center, The Netherlands

ARTICLE INFO

Keywords:
Psychometric models
Network theory
Dynamical systems
Measurement

ABSTRACT

In psychology, regulation of measurement practices relies on the use of psychometric models that relate the targeted psychological construct to a set of observations in empirical research. Recent work has argued for a new conceptualization, in which psychological constructs are viewed as networks. However, in contrast to existing models, network psychometrics has not articulated normative guidelines for measurement. To develop normative criteria for measurement in the network framework, a clearer picture is needed about possible measurement targets in a network. Using the example of depression, we explore five such potential measurement targets in the network framework: the *structure* of the network, the *current state* of the network, and three properties of the dynamical landscape, the *expected state*, *vulnerability*, and *stable state*. We discuss strategies for measuring these targets as well as challenges that arise in doing so.

1. Introduction

Many of the constructs that psychology is concerned with are not directly observable (e.g., disorders, cognitive abilities, and personality traits) but are assumed to manifest themselves indirectly in observable behaviors. Psychometric models are used to model the relations between these latent constructs and observable behaviors. Although deterministic models exist (e.g., [53,54], these relations are more commonly assumed to be noisy, both because there are errors in the measurement procedure (e.g., a respondent may accidentally cross out the answer next to the one intended) as well as because behaviors (e.g., 'going to parties' in the context of measuring 'extraversion') reflect myriad other factors than just the construct one intends to target for measurement (e.g., 'going to parties' might also reflect factors such as whether parents allow one to go to parties, or whether one is physically able to go). The relation between the construct and behaviors is therefore typically viewed as probabilistic, such that a higher level of the construct corresponds with a higher probability of a certain response. Psychometric models can be used to model the associations between item scores and to identify a variable from them so that scores on this variable can serve as a "measure of the construct". In contrast to other measurement theories, such as the classical [86] and representational theories of measurement [66], psychometric theories (sometimes called 'test theories'[1]) thus conceptualize the measurement problem *statistically* [77].

The construction of psychometric instruments characteristically proceeds by a process where a statistical model, which is deemed appropriate based on the construct theory, is fitted to the data, after which the test is adapted to the chosen measurement model (e.g., items are rewritten or removed to improve model fit) and the model is updated. An example may involve fitting a latent variable model as a statistical implementation of the hypothesis that individual differences on item scores on an intelligence test are determined by individual differences in general intelligence [107]. Items that do not fit the model, for instance because they reflect direct influences of background factors that are not mediated by the focal construct (i.e., item bias; [83]) may then be removed to improve the test, leading to a better measurement instrument. As a result, it is crucial that the model used is aligned with the construct theory: the substantive theory that characterizes the hypothesized structure of the construct as well as its relation to observations.

* Corresponding author at: University of Amsterdam, Nieuwe Achtergracht 129-B, 1018WS Amsterdam, The Netherlands.
E-mail addresses: r.vanbork@uva.nl (R. van Bork), g.lunansky@amsterdamumc.nl (G. Lunansky), d.borsboom@uva.nl (D. Borsboom).

[1] In this article, we choose to use the term 'psychometric theory' or 'measurement theory' instead of 'test theory' because test theory has a strong connotation of educational testing while measurement can be more general than the assessment of abilities. Classical test theory, generalizability theory and modern test theory are also entrenched with the term 'measurement': Lord & Novick [70] talk of "mental *measurement* theory", Cronbach et al.'s [28] generalizability theory classic is called "The dependability of behavioral *measurement*", and common terms in modern test theory are '*measurement* error' and 'reflective *measurement*'.

https://doi.org/10.1016/j.measurement.2024.114643
Received 15 February 2023; Received in revised form 22 March 2024; Accepted 31 March 2024
Available online 1 April 2024

Recently, it has been argued that, in many cases in psychology, the currently standard psychometric models are ill-suited as a match for many psychological constructs (e.g., [26,32]). The reason for this is that many psychological constructs are best conceptualized in terms of *systems* of interacting behaviors, affect states, and cognitions. Over the past decade, this realization has given rise to a new conceptualization of psychological constructs, in which constructs are viewed as *networks* of interacting components. For example, whereas in latent variable theory depression is conceptualized as a latent cause that manifests itself in observable symptoms, in network theory, depression is conceptualized as a network of symptoms that reinforce each other. This new analytic framework has become known as *network psychometrics* [78,80].

Although the development of network psychometrics has led to a plethora of new analytic techniques[11,42], which are now being applied to areas such as clinical psychology [96], attitude research [32], and intelligence [64], these techniques are primarily descriptive and have not yet resulted in an encompassing psychometric framework that can guide measurement. So, we need a measurement theory for networks. In the remainder of this article, we make some first steps towards developing such a theory. In taking this first step, we limit the scope to a specific type of psychological constructs: mental disorders. Throughout this manuscript we take depression as an example. We also focus on a specific family of network models that are used to model mental disorders: Ising models that represent symptom networks, in which the nodes are binary variables that represent symptoms that can be "active" (presence of the symptom) or "inactive" (absence of the symptom). This is not the *only*, and not even necessarily the *best* possible way of modeling a network of a mental disorder, however we take it as a starting point and will at several places discuss extensions beyond this type of networks.

We show that normative criteria about measurement are closely connected to how different psychometric theories conceptualize the measurement target, and argue that assumptions that underlie the norms of other psychometric theories, need not hold for network constructs.[2] To develop a measurement theory for networks, we need a better idea of what the measurement target is in network theory. We take depression as a leading example throughout the manuscript, and explore four possible measurement targets that naturally fit in with a network approach to mental disorders: the *structure* of the network, the *current state* of the network, and three properties of the dynamical landscape that characterizes the network, the *expected state*, *vulnerability*, and, the *stable state* someone is in.

The current article applies to the situation in which one has prior theory that the construct is best characterized as a network, and does not necessarily cover circumstances in which the network is used as a statistical method [11]; as such, this article is not a tutorial on how to deal with different types of variables or similar 'best practices' documents (cf. [59,19]). Neither do we intend to pass judgment on existing practices of test development or network analysis applications from a normative standpoint; rather, our interest lies in exploring possible measurement targets for network conceptualizations of mental disorders.

2. Psychometrics and the Network Framework

An important function of psychometric theories is to provide a normative framework for the measurement of psychological constructs that regulates test design and analysis [121]. Psychometric models are commonly used to distinguish between good or bad items, for instance by investigating them with respect to measurement precision [84], unidimensionality [108], and measurement invariance [85].

Different psychometric theories conceptualize the target of measurement in distinctly different ways; in classical test theory, the target of measurement is often thought of as an expected value over repeated administrations of the test, the so-called true-score [70], and in latent variable theory (also called 'modern test theory', which includes Item Response Theory [IRT] and Confirmatory Factor Analysis [CFA]) the target of measurement is conceptualized as a latent variable that acts as a common cause of the item scores [13,3,6]. A core assumption of the latent variable model is that scores on different items are conditionally independent given the latent variable (i.e., local independence), which is formally indistinguishable from the causal condition that a common cause screens off correlations between its effects. Do note that while the common cause hypothesis can justify the latent variable model [113], and has intuitive appeal [8], it is not necessary for the application of the model; one can in fact derive the model from different starting assumptions (see, for example, [39]). However, even under a pragmatic or instrumentalist interpretation of the latent variable model [60], a researcher who uses this model should be minimally prepared to defend the assumption that the observational data satisfy certain conditions that would follow naturally from a common cause model.

To compare the different conceptualizations of the measurement target in classical test theory and latent variable theory in a concrete example, consider the items of Beck's depression inventory [2], which includes questions on behaviors and feelings like how irritated the person feels, whether the person cries more than usual, and whether the person sleeps as well as usual. In latent variable theory the property that is targeted by asking such questions is conceptualized as a latent variable that causes these behaviors and feelings, and in true-score theory the measurement target would be someone's expected value of the total score on this questionnaire over hypothetical repeated administrations.

The norms that accompany these measurement theories also depend on these different conceptualizations of the measurement target. For example, what part of the variance in item scores is considered *reliable* variance versus *error* variance differs between classical test theory and latent variable theory.[3] Similarly, the use of a unidimensional latent variable model lays down as normative criteria that item scores are conditionally independent given a fixed level of the latent variable (no residual correlations),[4] that items should have adequately high factor loadings, and that the error variables should not differ in means or variances between social groups (such differences would entail group bias). If the measurement target is a common cause of the item responses, so that the construct theory matches the psychometric model, then selecting items based on these criteria indeed improves measurement of the targeted construct. Alternatively, use of the true-score model [70] lays down as a normative principle that reliability should be sufficiently high, which, if Cronbach's alpha is used as a criterion, means either that items should have strong intercorrelations or that many items

[2] Note that we use the term 'construct' more broadly than the commonly used notion of an 'attribute', which is typically used to denote a single dimension of interest. This allows us to include the possibility of having a network (i.e., a high-dimensional structure of interactions between variables) as a construct.

[3] In classical test theory, the variability between people's expected values is reliable variance and deviations of observed scores from these expected values is error, which is consistent with the target of measurement being conceptualized as an expected value. In modern test theory, variance in the item scores that is shared with the common factor is reliable, and the variance in item scores that is specific to the item or a small subset of items is error variance, which is consistent with the target of measurement being conceptualized as a common cause. Only under certain assumptions do these two ways of distinguishing between reliable and error variance give the same results. Note, however, that not all modern test theory models redirect all unique variance to the error term: there are models that further decompose unique variance into random measurement error and item-specific factors (see for example, [103,110]; and see [112], for more details on interpretations of error in classical and modern test theory).

[4] There are less restricted latent variable models that do allow for some residual correlations, but generally residual correlations are unwanted as they violate unidimensionality which is considered important for the interpretation of test scores [37].

should be used [106]. Because reliability is the squared correlation between true-scores and observed scores, selecting items based on Cronbach's alpha as an estimate of reliability makes sense if the target of measurement is conceptualized as a true-score variable (an expected value over repeated administrations).

Note that although criteria are derived in a certain psychometric framework, they need not be exclusive to that framework. For example, reliability is more strongly emphasized in classical test theory than in latent trait theory, but is also of importance in the latter [84]. Also, the fact that certain criteria are derived within a psychometric framework (e.g., reliability in classical test theory) does not mean that there is a strict order where there is first a psychometric theory, and then normative criteria fall out of it. Normative criteria might well shape psychometric theory in turn. Our point here is that normative criteria fit with a conceptualization of the measurement target so that large shifts in conceptualizing the measurement target ask for reconsideration and possible adjustment of these criteria.[5]

One such example of a large shift in conceptualization is offered by the network theory of mental disorders. In the next section, we will take a closer look at a network theory of mental disorders, before we move to possible measurement targets in that context. Although the term 'symptoms' for problems associated with a disorder is often interpreted as implying that these problems are caused by the disorder, we will use the term 'symptom' to refer to a broader meaning in which these problems are not necessarily *caused* by the disorder and use it also in the context of the network framework.

3. Symptom Networks as Ecosystems

An important analogy that guides thinking in network approaches to psychological constructs is the analogy of an ecosystem [116,71]. Just like correlations between measures of the quality of water and biodiversity in the ecosystem of a lake can be explained through mutual reinforcement without introducing a common cause like 'lake health' to explain away the positive correlations, the correlation structure observed in psychometric analyses of symptom sets could also be explained without a common cause: causal interactions between symptoms explain the correlations between measures of these symptoms.[6] Thus, in this view, symptoms are not passive indicators of a latent attribute, like a 'central psychological process' [67] or a 'brain circuit disorder' [57]. Rather, they are active causal agents that mutually reinforce each other. For example, in psychopathology, symptoms have been proposed to be mutually reinforcing components of a *symptom network* [9,12]. In such a network, symptoms are represented as binary nodes that can be active ("presence of symptom") or inactive ("absence of symptom"), and the interactions between symptoms are represented by edges that connect these nodes.

In theory, interactions between symptoms in a tightly connected symptom network can lead the network to 'get stuck' in a state of high symptom activation [25]. This can happen if the interactions between symptoms are so strong that the symptom network is capable of sustaining its own activation [9,12]. As such, much like an ecosystem of a lake can show sudden transitions between a turbid state and clear water state [98,116], symptom networks could suddenly shift between a healthy state with low symptom activation and an unhealthy state with high symptom activation. Using very simple models that are based on the Ising model from physics [58], simulations of such networks have been shown to display hysteresis [25]: In these simulations, networks are subjected to increasing external stress (operationalized as an external field that increases the chance of symptom activation). Such simulations show that the level of stress at which a strongly connected symptom network transitions from a healthy state into a stable state of high symptom activation is higher than the level at which it returns to the healthy state. This phenomenon is known as hysteresis and implies that an external perturbation may lead the network to become activated, but removing that perturbation may not lead to a return to the healthy state.

For ecosystems, the dynamical landscape of the system is typically illustrated using the ball-and-cup metaphor (e.g., see [101]). A picture of the ball-and-cup metaphor can be found in Fig. 1, and represents a theoretical stability landscape in which the horizontal axis represents states with different node activation, and the depth and steepness of the 'cups' represent the stabilities of such states, so that two cups as in the lower picture can represent a healthy stable state (low symptom activation) and a depressed stable state (high symptom activation). It is possible to make the surface on which the ball moves multidimensional, so that each different combination of active and inactive nodes has a stability outcome, but often the location on the horizontal axis represents a summary state, such as the total number of active nodes. The steeper and deeper the cup, the more likely the ball will end up in the state at the lowest point of the cup and the more stable that state is. Network approaches to psychopathology have used the same ball-and-cup metaphor, as represented in Fig. 1, to describe the difference between resilient and vulnerable individuals. The hypothesis is that, for resilient individuals, the parameters of the network structure are such that the system will return to its healthy state upon perturbation (Fig. 1, top panel). For instance, with respect to the symptoms of depression (including sad mood, lack of appetite, loss of interest), the loss of a loved one is an external stressor that may temporarily induce an elevated symptom state that superficially is similar to clinical depression. However, as time passes, and the impact of the external stressor wanes, the system will return to baseline. However, for a vulnerable individual, the parameters of the network are such that, upon a perturbation, the system may tip into another stable state, which is characterized by persistent symptom activation (Fig. 1, bottom panel). This stable state is what the clinician perceives as the phenomenology of the mental disorder in question. Simulations have shown that networks with low thresholds for the nodes (so that lower levels of external stress can already activate the nodes) in combination with strong connections between nodes (so that the activation of nodes is more strongly dependent on the activation of other nodes) are characterized by more vulnerability to end up in a depressed stable state [71].

The network approach, as outlined above, is a generic framework to understand psychological constructs as systems of interacting elements. As such, it is just systems thinking applied to psychological constructs. However, one of the reasons that the network approach has become popular in empirical research is that it has been complemented by a range of statistical modeling approaches that can be used to estimate networks from empirical data [11,41,42,79]. These models were implemented using regularized estimation techniques that were originally developed to estimate Ising models in high dimensions [114,49,82]. Since, a plethora of techniques have been developed to estimate networks using various kinds of data, different estimation methods, and different goals [11,78,80].

A review of these techniques is beyond the scope of the present article, but it is important to briefly discuss how the typical models used in the literature work. Characteristically, the models represent variables as nodes, and conditional associations as links between nodes. This is essentially the same representation as is used in the tradition of

[5] Note that we do not claim that the criteria implied by assumptions in psychometric models are the only criteria that are used for test development, nor that they are most important. For example, an important other criterion for test development is validity. Different theories of validity have been developed [77,92] and these have resulted in several criteria that are also applied in test development (e.g., test-criterion correlations, measures of convergent and divergent validity).

[6] Note that symptoms are not identical to symptom reports (i.e., recorded entries in a dataset); for instance, if a symptom is assessed with multiple items, it may be represented by a diagnostic cutoff on the sum score on these items, by the sum score itself, or by a latent variable (see [27,45]).

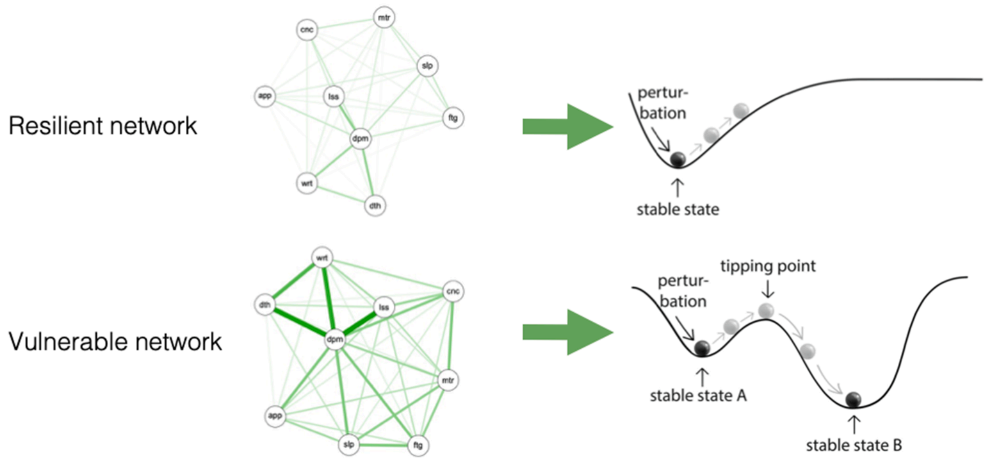

Fig. 1. *Ball-and-Cup Metaphor for resilience* versus *vulnerability. Note.* A resilient symptom network will return to a healthy stable state upon perturbation (top panel). In contrast, a vulnerable symptom network may enter an alternative stable state of persistent symptom activation.

graphical models (e.g., see [24]). The idea is that associations between variables that represent components that causally interact cannot be fully explained away by conditioning on the remaining variables in the data, and hence should remain present under conditioning, while associations between nodes that result from a shared dependence on the remainder of the network should vanish under conditioning. As such, the statistical approach offers an exploratory way of assessing the network architecture of psychometric constructs; namely, by running the relevant models on the responses to a set of items [11].

Although this approach has become known as network psychometrics, it should be noted that the network model is merely a statistical model. Such a model can represent the conditional associations that are present in the data, but it is not a psychometric measurement model in the sense that it uses the observed scores to estimate a person's position on a targeted construct. Network models currently do not provide normative criteria for what items to select and how to combine the information in item scores into a measure of the construct.

Because in the network framework constructs are modeled as systems of interacting nodes (e.g., behaviors, attitudes or symptoms), the psychometric criteria that we discussed in the previous section that follow from classical test theory or latent variable theory do not automatically transfer to a measurement theory for networks. For example, in the network framework, there is actually no reason to expect that local independence would hold in principle (although it is not impossible either; see [44,79]). Similarly, while variance in the item scores due to item-specific causes is considered 'error' in the latent variable model [112], unique variance in item scores is not necessarily considered error in the network framework, and there is no reason to only consider the shared variance between items. In fact, applying norms of a measurement theory that does not align with one's theory about the construct can do more harm than good. If the target of measurement is *not* the common cause of the behaviors that are measured with a set of items, but instead, for example, the effect[7] of these behaviors, the procedure of removing item-specific variance to deal with measurement error can induce bias [95]. This illustrates the importance of using a measurement theory that is consistent with how the target construct is conceptualized.

4. Measurement Targets in the Network Framework

The first part of this article has pointed out that existing measurement theories include normative criteria that are inconsistent with network theories of psychological constructs. Measurement of psychological constructs in the network framework thus requires the development of new or adjusted measurement theory that fits with a network conceptualization of the construct. A first step in doing so, is to flesh out what the target of measurement is in the network framework. Taking the construct 'depression' as an example, whether some measurement *X* is a good measure of depression, will depend on what one tries to target with a measure of depression. In line with the traditional measurement of "attributes" in psychometrics, and measurement of "properties" in metrology [75], the measurement targets we discuss refer to properties and not to the objects that bear the property. There is no agreement in the literature on whether nominal properties can be 'measured' or whether alternative terms should be used for examining such non-quantitative properties (see e.g., [73–75,86]). We choose to use a more tolerant definition of measurement that allows for the measurement of quantitative, ordinal and nominal properties, in line with Mari et al. [75]. In the following, we identify five important potential measurement targets for the example of depression in the network framework. These five targets are (1) the network structure, (2) the number of active nodes at the moment of measurement (current state), and three properties of the dynamical landscape: (3) the number of active nodes that is expected from the dynamical landscape (expected state), (4) the ease with which someone moves into a depressed versus healthy state after perturbations (vulnerability), and (5) whether someone is currently in a healthy or depressed stable state. Fig. 2 shows a visualization of these five targets. Target 1 is a nominal property, targets 2, 3 and 4 can be ordered such that a lower score is healthier, and target 5 is a binary property (healthy/depressed). For each of these, we first describe the target and then discuss considerations for measurement of these targets.

The relation between the conceptualization and measurement is not unidirectional; not only should theory about the construct inform measurement, the way we go about measuring a construct will also

[7] A model in which the construct is a common effect of the indicators instead of a common cause of the indicators, is known in the literature as a *causal indicator model* [4].

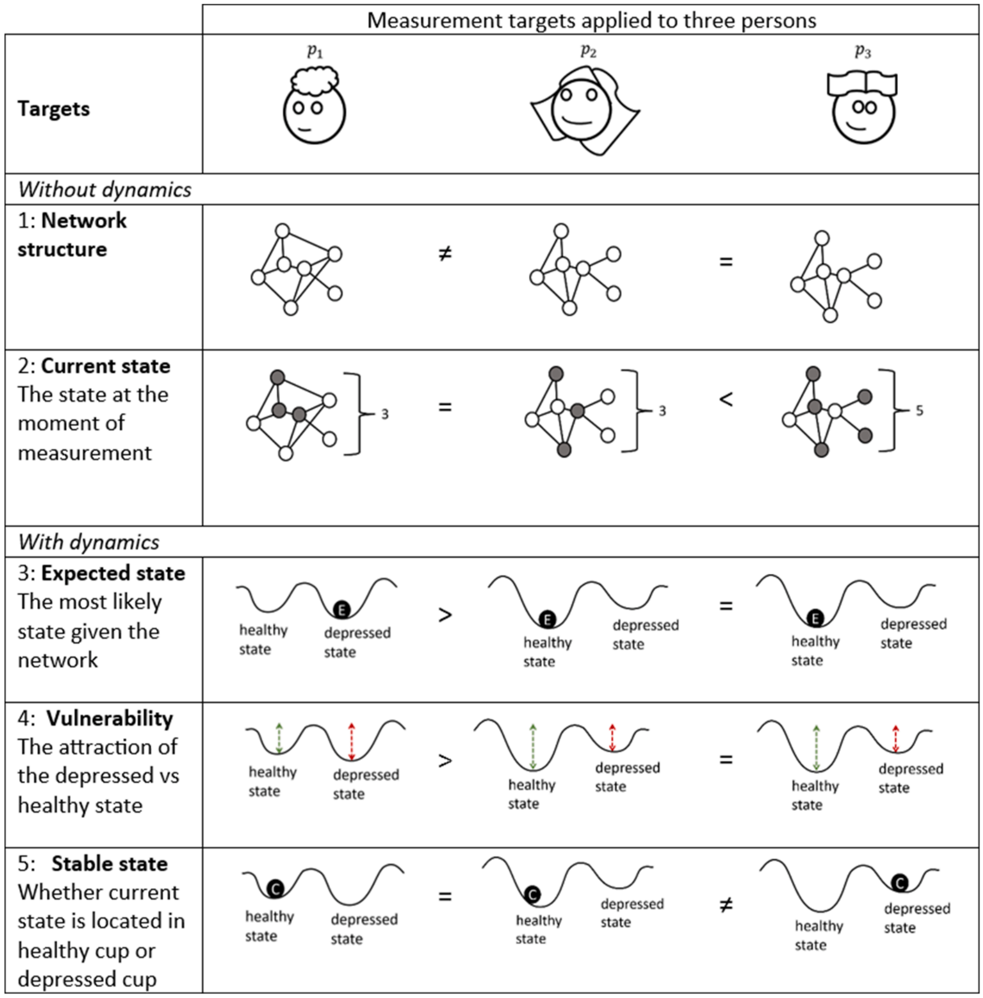

Fig. 2. *Visualization of Possible Measurement Targets in the Network Framework. Note.* Targets 3, 4, and 5 are depicted with a ball-and-cup analogy similar to that in Fig. 1.The ball of the stable state has a "c" in it, because it represents where the ball currently is (current state), whereas the ball of the expected state has an "e" in it, because it represents where the ball most likely is given the landscape. For individuals with the *same* network structure, when a dynamic is added to their networks, they will have the same dynamical landscape and as such share the same expected state and vulnerability (persons p_2 and p_3). However, they can differ in current state, and in stable state, as these two targets concern what the state is at the moment of measurement.

shape how the construct is conceptualized[8] [20]. We therefore explore different properties not to choose *the* measurement target as something that should be settled before going into measurement. Rather, we explore these measurement targets from the view that more clarity on possible conceptualizations of a construct can aid the development of measures for this construct, and we expect that future developments in measurement of network constructs can further sharpen these conceptualizations as well as develop measurement theory for these targets. In the following, for each of the five targets (network structure, current state, expected state, landscape vulnerability, and stable state), we first discuss what this target is and why it might be an interesting measurement target, and then, in the subsection "Considerations for Measuring [target]", discuss what has been done so far to measure these targets and some of the challenges that come into play. In reading through the challenges that are raised with the different measurement targets it might be helpful to recognize a certain hierarchical structure among the targets. Some of the measurement targets sit on top of other measurement targets so that any challenges that apply to one target are automatically also a challenge for the other target: Measurement of the network structure relies on measurement of nodes and measurement of edges, and so any challenges in measuring the nodes or edges will also be a challenge for measuring the network structure. The expected state, landscape vulnerability and stable state are properties that require measurement of the network structure so that any challenges in measuring the network structure apply also to measurement of these targets.

4.1. Targeting the Network Structure

One possible target for a measure of depression is to measure the network structure. That is, 'depression' is conceptualized as a set of depression symptoms and causal relations that connect these symptoms and we want to represent the structure of this network as closely as possible by representing the symptoms as nodes and representing the causal relations between symptoms as edges between nodes. To do so, one could measure the symptoms and estimate which symptoms are causally related. Depression would then be measured well if the

[8] As an example, viewing depression as a dynamical system is one factor that contributed to the use of Experience Sampling Method (ESM) in which subjects monitor their subjective experiences (e.g., different emotions) over multiple consecutive days, and which provides the type of time-series data that could be used for predicting a state shift into a depressed state (e.g, [115],[117]). This choice of measurement in its turn shapes the conceptualization of depression, for example by giving a prominent role to mood or other variables that fluctuate sufficiently over time and a focus on the significance of state shifts.

estimated network structure has edges between any two symptoms A and B, if A causes B, B causes A, or both cause each other, and no edges between symptoms for which neither of this holds. The structure of the network can thus be seen as a property of the network that refers to a qualitative pattern that can be characterized as a set of nodes and edges, and possibly additional information on these nodes and edges such as the strength of these edges or thresholds of the nodes. There are also properties of the network that can be targeted that summarize certain information of the network structure, such as the mean edge weight, number of edges, or proportion positive to negative edges, etcetera. Here, we focus on measuring the set of nodes and its connections.

An important difference with the psychometric theories described in the first section is that by taking the structure of edges that connect nodes as a measurement target, it is explicitly not a single dimension or line on which one can order people. Either, one targets a population network structure that represents relations between individual differences in symptoms, or, one can target individual network structures (each person having a 'depression network' which includes the relations between symptoms for an individual over time). Estimating such a network from cross-sectional data is possible but requires assumptions on either the homogeneity [46] or the distribution of network structures over individuals [80,78] that may not be satisfied [88]. In addition, if there is heterogeneity in network structures, the assessment of individual differences in that system can be an important goal [46,88,51,72]. For instance, to investigate heterogeneity one may use idiographic research (study of a single person over time; e.g., see [56]) to estimate individual networks from time-series data, which typically requires other assumptions such as stationarity. Alternatively, one can assume a distribution of network structures in the population as a Bayesian prior, and use this to arrive at a posterior over these networks given a person's item responses [80,78].

Importantly, apart from the challenges with estimating individual networks, even if we would have access to the networks of individual people, it is not trivial how one would order these networks in a way that it becomes possible to say that one person is more depressed than another. That is, the network structure as a target either is a property of a population, or in case of individual networks, a property of a person, but in both cases it should be considered a nominal property. We can identify *differences* between networks (e.g., person 1 has a different network than person 2, or person 1 has a different network before treatment than after treatment), but not *order* them (see Fig. 2).

4.1.1. *Considerations for Measuring the Network Structure*

Three challenges in measuring the network structure as described above are the selection of nodes, the selection of items that measure the nodes, and the selection of edges to connect the nodes. Below, we discuss each of these in their own subsection. The next four targets will add additional challenges but will also have to deal with these challenges.

Selecting Nodes. We chose to focus on symptom networks, but how to decide whether something counts as a symptom of depression? And why include only symptoms and not also other factors that causally impact the network (see e.g., [61])? And, in case other relevant causal factors are included, where does the network stop [90]? Even in the case of choosing to include only symptoms, it will be difficult to demarcate depression from a larger network of mental health problems in which it is embedded (e.g., other symptoms of other mental disorders that connect to depression symptoms). A significant challenge in targeting the network structure will thus concern the selection of nodes.

Often, the nodes that are included in a network are taken to coincide with a set of items in a test or questionnaire, such that each item is a node. But, we should actually distinguish two levels at which selection takes place: the selection of nodes to best represent the system one tries to target, and the selection of items to measure the nodes. As such, the nodes are lower-order targets of measurement that are needed to measure the higher-order depression construct. We will first discuss some of the considerations in selecting nodes, to then discuss possible normative criteria for items that should measure these nodes.

There are two main challenges in selecting nodes. First, one has to *demarcate* what counts as part of the network of depression and what does not [90]. Several approaches have been pursued in the literature in this respect. For example, in the case of depression, existing symptom sets that have traditionally been taken to define depression according to the DSM-5 have been used as nodes in a network [96]. In addition, whatever items happen to be included in a questionnaire have been incorporated in network analysis (e.g., all items in the Beck Depression Inventory; [17]). Further, nodes have been selected based on expert opinion concerning which factors are important in depression [122]. Finally, nodes have been selected on the basis of self-reports, where patients themselves identify factors that are important in their case [69]. The selection processes have thus been quite varied; sometimes nodes have been selected on purely pragmatic grounds (e.g., in secondary data analysis) and sometimes using an intensive procedure (e.g., expert interviews).

Networks of psychopathology typically use symptoms or momentary states such as thoughts and affective states (mostly in networks of time series data) as nodes, but this is not theoretically necessary[61,16]. Note, however, that factors that are *not* included in the network of a construct, are not assumed to be unimportant; they can still affect nodes in the network, but would do so as part of the "external field" (a term borrowed from the Ising model in physics, used to denote factors that affect nodes in a network from outside; [9,12]). For example, stressful life events will have an influence on depression symptoms, but are arguably not "part of" depression as a system.

Suppose one settles for the inclusion of only depression symptoms in the network, a second challenge then amounts to coming up with criteria for which symptoms are important to include in a network, and which are less important. Maybe for symptoms of depression it could be possible to argue that 'all' symptoms should be included. But, for example, for a network of a personality trait or an attitude, there is typically no exhaustive list of behaviors that can all be included [90]. A possible criterion could then, for example, be that nodes are selected such that the network of the selected nodes captures important characteristics of the underlying total network. This, as well as the demarcation of what is and what is not part of the network of a certain construct, are still open questions.

If the construct is a causal network, then it becomes important that the nodes included in the network are distinct autonomous causal components [21,22,33]. For example, items that measure the same property should not be included as separate nodes in the network. Christensen, Golino, & Silvia [22] give the example of items such as 'hate being the center of attention', 'like to attract attention' and 'dislike being the center of attention' (all items of the SAPA personality inventory; [23]), for which you would rather include "attention seeking" as a node in a network of extraversion, than creating nodes for all of the separate items that measure this property. One normative criterion for the selection of nodes could thus be that the nodes are *separately identifiable*[9] [16], which echoes the older idea that causes and effects should be distinct, as articulated for instance in Hume's maxim that "an effect is a distinct event from its cause"[104].

Selecting Items. In addition to criteria for the selection of nodes, there can be criteria for the selection of items to measure these nodes. Christensen, Golino & Silvia [22], for example, write: "When evaluating whether our questionnaire is a valid measure of extraversion from the

[9] Separate identifiability does not imply conditional independence. It is a requirement that comes from theories of causality, where a standard criterion is that causes and effects should be separately identifiable. For instance, "being a bachelor" cannot cause "being unmarried", because there is no way of separately identifying these properties. It is thus possible for two variables to be separately identifiable but not conditionally independent in the population (e.g., smoking and lung cancer).

network perspective, we must shift the evaluation from the validity of extraversion as an attribute to the validity of its components." (p.1098) Given that the nodes in the network represent such distinct autonomous causal components, this thus means a shift to measurement at the level of the nodes. For example, for each of the nodes, one could ask whether the item scores (e.g., responses to the question "how well did you sleep last night") are good measures of the intended node (e.g., 'sleeping problems'). That is, one could introduce a small latent-variable or true-score model for each individual node [27]. Currently, nodes are typically measured with a single item, so that the network consists of as many nodes as items in the test. But, it is of course possible to measure a node with multiple items, and, for example, include a latent variable model to estimate node scores from multiple items [45,36].

Although the network theory might conflict with the use of the latent variable model at the level of measuring the higher-order construct depression, this might be different for the measurement of nodes. For example, for two items, one asking about sleep problems and one asking about concentration problems, a latent variable model of depression explains the correlation by both of these items reflecting their common cause 'depression' (having sleeping problems is indicative of the person being more depressed and people who are more depressed tend to also have concentration problems). Instead, a network theory of depression would view these two items as measuring nodes that are distinct causal components that reinforce each other (e.g., because someone sleeps badly, the person is tired and has a hard time concentrating). Yet, for two different items, one asking the amount of agreement with the statement "I had a good night of sleep" and one with the statement "I laid awake for a long time", the (negative) correlation would also in a network theory of depression be more naturally explained by viewing these items as reflective of the same component as opposed to measuring two distinct components that causally interact. For such items it might be better to either only include one of them, or, even better, use the combination as a more reliable measure of the node 'sleeping problems', using, for example, a latent variable model, congeneric true-score model or sum score. In that case, normative criteria from other psychometric theories can be used for measurement at the level of the nodes.

What psychometric theory can best be used for measurement at the node level will rely on how the nodes are conceptualized. That is, similar to the question of whether the higher-order construct 'depression' can best be understood as a latent common cause, expected value over repeated measures, composite, network or other, one can also ask this about the symptoms that are represented by the nodes. Depending on the answer to this question one can choose a measurement model that best fits this conceptualization. Although it might not be straightforward to choose the appropriate measurement model for the nodes, our main point here is that even when a higher-order construct is conceptualized as a network and thus measurement of this construct would be inconsistent with the use of latent variable or true-score theory, this does not mean that the nodes cannot be measured with latent variable, true-score or other traditional psychometric models. We think it would be good if more thought goes into how to best measure nodes (see e.g., [16] for a discussion on 'node validity'), and in which cases it would be better to replace the single-item measures of nodes that are now common practice, with measurement models at the node level.

Selecting Edges. Given a specified set of nodes, the question arises how to decide which nodes should be connected with an edge. Again, several approaches have been followed in this respect. First, edges have been taken to exist on the basis of theory; e.g., nodes have been connected whenever symptoms occur in the same disorder (a bipartite network; [10,111]). Alternatively, edges have been included on the basis of self reports, where individuals are asked, for each combination of symptoms, whether one symptom causes the other (Perceived Causal Network approaches; [65,50]). In addition, edges have been included on the basis of expert opinion, where experts are asked to rate the extent to which symptoms causally interact [35,122]. In yet another approach, edges have been included on the basis of time series data, where individuals are tracked for a longer period of time and answer small questionnaires multiple times a day (experience sampling methodology); vector autoregressive models are then used to assess conditional associations between symptoms over time [51,18,56]. Extending the network approach to directed networks, causal search algorithms have been used to exploratively search for causal network structures [81]. Finally, cross-sectional data have been used to assess conditional independence relations that characterize the population distribution of symptoms, where nonzero conditional independencies define weighted edges in the network [114]; this approach has, in the past few years, been most prevalent [96], probably because of the wide availability of this kind of data and the relative ease with which one can estimate a network for it.

If one's objective is to assess a person's network structure, then each of these approaches could be viewed as an alternative way to assess the same set of edges. One could therefore think of the assessments of an edge as being measures of the same property (e.g., the strength of the edge) through different means. In cases where one has assessments of the strength of the same edge in different individuals, one could conceptualize differences in edge strength as a latent variable and use classic psychometric models to assess reliability and validity.

One could say that the network is measured well if the network structure (presence and absence of edges) matches the true relations between those variables included as nodes in the network. For example, one could say that the estimated network is a good measure of the "correlation network", if the edges correspond with nonzero correlations, and is a good measure of the "causal network", if the edges correspond with causal relations. Because networks are high dimensional structures, the degree of correspondence is not easily captured in a single number; statistics that are commonly used to assess network recovery include the probability that a procedure selects connections that actually exist (sensitivity), the probability that nonexistent connections are correctly set to zero (specificity), the correlation between estimated and true edge weights, and a number of other statistics that assess correspondence, such as Kendall's tau and the Jaccard index [114,48,12,9,43].

4.2. Targeting the Current State

The previous target was a nominal variable on which people can differ but can not be ordered on *how* depressed they are. A straightforward way of assessing how depressed an individual is, is by counting the number of nodes that are active in the network at that time point. In case of a depression network with symptoms as nodes, this amounts to counting the number of symptoms that are currently present. The number of active nodes can be called the *state* of the network [71]. Some literature uses the word 'macrostate' for this term, to distinguish it from 'microstates' which refers to all of the separate possible combinations of active and inactive nodes [30]. So, for two people who have five nodes that are active in their depression network (five symptoms present), their microstates are different when they have different symptoms, but they will have the same macrostate (five active nodes). We will refer to the macrostate of the network at the moment of measurement as the 'current state', which can be targeted for different reasons. Most importantly, the number of active nodes will often provide a useful summary of the overall condition of the system.

If the number of active nodes is targeted as a variable that represents depression severity, then an implication of choosing this target is that it is the number and not the nature of the symptoms that counts towards severity. In a network, different nodes can play very different roles; some might vary in line with their neighboring nodes, while others vary more independently of the rest of the system. Some might mostly cause other nodes, while others are mostly caused by the other nodes. Such information on the role of a node in the larger system is not accounted for in the number of active nodes. In fact, one does not need to know anything about the network structure to count the number of active nodes. This will start to differ when the target is not the current state of the measured network, but the current state of the 'full' network, of which only some nodes were measured to estimate a network. For example, suppose that we assume that the set of nodes in the estimated network is a sample of nodes from some true network. In that case, an alternative target could be the current state of the full network. We can thus distinguish between measuring the current state of a network for a fixed set of nodes, and the current state of a full network from which only a sample of nodes has been measured.

4.2.1. Considerations for Measuring the Current State

The current state of a network is basically the sum score of the nodes that are active in the network at the time of measurement. To measure this sum score one needs to select the nodes of the network, and select items that measure the nodes (two challenges discussed in the previous section), but, in contrast to the previous target, one does not need to know the edges between the nodes if one can assume that all nodes in the target network are measured. If all nodes are measured, and each node is measured with a single binary item (which is not unusual in symptom networks), the current state can be measured by taking the sum score of the item scores.

If one assumes that there is substantial measurement error involved in measuring the nodes, one can use multiple items for a node and, for example, use a congeneric model to estimate a true-score for the node. In that case, although the number of active nodes is an integer, the resulting score, which we will call the 'node score', that is used as a measure of the number of active nodes, no longer needs to be an integer, as it uses expected values for the nodes.

The sum of active nodes can be viewed from various angles. First, the node scores can be seen as *formative* indicators of the macrostate [5,38,4,76], in the sense that the macrostate is a function of the node states rather than the other way round. If not all nodes are observable, a set of nodes that is deemed representative can be selected. The sum score of the selected set of nodes could be used as a measure of the total level of activation in the network as a whole, similar to the definition and use of index variables in fields that commonly assess the overall state of a system, such as the use of indexes in economics (e.g. the Dow Jones as an overall indicator of the state of the U.S. economy).

If the target of measurement is the current state of a fixed set of nodes, this is similar to how items are typically treated as fixed in latent trait theory and classical test theory [34]. However, if the selected nodes are treated as a random sample of the total set of nodes of the underlying system, the nodes are considered random, which brings it closer to domain sampling theory in which items are considered a random sample from a population of items. In that case, the activation in the network as a whole can be construed as a universe score as in generalizability theory [28,15].

From the point of view of network psychometrics, one may ask to what extent the node score of a network of a selection of nodes suffices to characterize the current state of the network as a whole. The *number* of active nodes among nodes in the estimated network will not be a good measure because the full network has a different total number of nodes. Instead, one can take the mean activity across nodes in the estimated network as a measure of the mean activity across nodes in the full network. This comes close to the notion of a mean field, as used in statistical physics. Mean field approximation essentially uses the mean number of active nodes (the mean field) to describe network behavior [47]. Interestingly, mean field approximations work especially well if the network is fully connected with invariant weights, which defines a special case of the Ising [58] model that is known as the Curie-Weiss model [62]. For a fully connected network with invariant weights, the mean field better approximates effects of individual interactions between nodes (the more similar the edges, the more exchangeable the nodes are, and the better the average functions as a stand-in for groups of nodes). For such a network, one therefore expects the mean activity of a random sample of nodes to better approximate the mean activity of the full network.

Interestingly, the Curie-Weiss model turns out to be statistically equivalent to the Rasch model [79], a paradigmatic example of a latent trait model for which the total score is a sufficient statistic [1]. Even more so, the fact that the total score is a sufficient statistic for items that meet the assumptions of the Rasch model, is used as a justification for the selection of items based on the Rasch model because it supports the additive decomposition of item response probabilities into person ability and item difficulty[94,7]. In a similar line of reasoning, one could take fit of a Curie-Weiss model as a normative criterion to select network items so that one ends up with a network for which the node sum score is a good measure of the current state. However, while this might be reasonable for some constructs, we think this should not be a general normative criterion because for many constructs we expect that important components in the network do not meet these constraints in which case this criterion would lead to sets of nodes that are not representative of the targeted network. That is, if the targeted full network itself does not follow a Curie-Weiss structure, then selecting nodes based on that criterion will result in a network for which the sum score might be a good measure of the current state of that selected network, but not of the targeted network.

In a network with a nontrivial topology, where nodes cannot be assumed to be interchangeable, some of the measured nodes could be more informative of the current state of the network as a whole than others. In that case, one would want to weight nodes by how informative they are about the current state of the full network. A possible candidate for weighting nodes is, for example, one of the node centrality indices (degree, betweenness, or strength; see [14], for an example). It is also important to note that the network state may not allow characterization by a single value; this may for instance be the case if the network displays strong clustering so that the characterization of its current state is better represented in a multidimensional way [44]. Christensen & Golino [21] introduced 'network loadings' as an equivalent to factor loadings, that represent a node's strength for a cluster of items in the network that would represent a dimension. Each node can thus have multiple network loadings for the different clusters in the network, and in such cases the network state would be better characterized by a vector than by a scalar value.

At this point there is no justification spelled out for the weighting of nodes on the basis of network structure. This is not to say that such justification could not be developed. For instance, a node with higher network loadings may be more informative than nodes with lower network loadings, because the strength a node has in the observed cluster is indicative of the strength of the node in the full network. And a node that has a high strength in the full network may be more informative with respect to the state of other nodes in the full network that we did not observe, than a node that is only weakly connected which indicates that it varies more independently of other nodes in the full network. However, although such reasoning might sound plausible, extensive simulations or derivations are needed to evaluate which properties of a node are indicative of the current state of a full network from which this node is sampled, and under which conditions. For example, strength typically is computed from the conditional relations between nodes. But, simulations may show that, as a predictor of the current state of the full network, the strength of simple correlations of a node with other nodes performs better than the strength of conditional correlations. There are also centrality metrics that are specifically constructed to represent node predictability [55], which may perform better at predicting the state of a full network.

When weighting nodes by any such metric (e.g., network loadings, centrality indices, covariation with other nodes), another interesting question pops up: what if the networks, and therefore these node properties, substantially differ between groups, or individuals? This question relates to the traditional notion of measurement invariance [83,85]. In the case of a factor model, a set of items is strictly measurement invariant if and only if the factor loadings, error variances and intercepts are equal across groups. If measurement invariance is violated, for example, because items differ in loadings across groups, then you cannot meaningfully compare groups on their factor scores. Similarly, it might be that for meaningful comparisons of node scores between groups or between people, the node weights (e.g., network loadings or other) that are used to get a node score should be invariant over groups. But, to draw such conclusions, simulations, derivations or other investigations are needed of what the consequences are of violating such invariance. Similarly, further investigation is needed into the conditions under which the sum score of an estimated network will be a good measure of the current state of a full network and in what cases the sum score will start to fail approximating this measurement target. For now, we recommend caution in cases where the network structure clearly diverges from a Curie-Weiss model.

4.3. Targeting the Expected State

As described previously, much of the theory in the network framework draws from dynamical systems theory. Using the ball-and-cup metaphor, the current state only considered where the ball is at the time of measurement, but ignored the dynamical landscape in which the ball moves. As an alternative, one can ignore where the ball is at a certain moment in time but instead consider how healthy the landscape is. For example, given the landscape, what is the most likely state for the ball to be in? And, given the landscape, does the ball easily roll back to a healthy state after negative perturbations? The dynamical landscape is a property of both a network structure (e.g., changing the thresholds or edge weights will change the dynamical landscape) as well as some dynamic process that specifies the development of node activation over time. As such, the dynamical landscape could be seen as the dynamic counterpart of the network structure. In the following, we discuss properties of the dynamical landscape, starting with the expected state.

It is possible that two people who have the same symptoms and their networks are thus in the same current state, differ in the number of symptoms that are expected to be active based on their different networks. For example, if for one of the persons the network is such that all symptoms are strongly connected to each other, and the other has a weakly connected network, then it is more likely for the person with the strongly connected network that other symptoms activate as well (this will also depend on the thresholds; [71]). One could therefore argue that for two people with the same number of symptoms at a certain time point, but different network structures, the person who has a network structure that is characterized by a higher expected value of active symptoms over the different states that result from the network with its dynamics, is more severely depressed. An alternative dimension that could be targeted for measurement is therefore the mean number of active nodes that is expected based on the network and its dynamics, which is called the 'expected state' of the network which differs from the 'current state' (the number of active symptoms at the moment of measurement). This target is visualized in the third row of Fig. 2. Just like for the current state, also for the expected state one can target the expected state of the nodes that are measured, or target the expected state of a larger unobserved network of which the measured nodes are only a subset of all nodes.

4.3.1. Considerations for Measuring the Expected State

The expected state is a property of the network that depends on the dynamics that characterize network evolution. So, whereas for a fixed set of nodes one does not need to know the network structure to measure the current state (one can sum up the nodes that are active without knowing anything about edge weights or thresholds), for the expected state one does need to know the network parameters. However, this also means that whereas for the current state it is possible to have individual differences in numbers of active nodes while people share the same network structure, individual differences in expected states *require* different network structures for different people. After all, two people with the same network structure will have the same expected state when you add a dynamic process to both networks. This means that for measuring people's expected states, individual networks need to be estimated which requires approaches that are able to represent structural differences between people in the model. Also note that because the network structure is needed to estimate the expected state, all of the challenges of measuring the network structure that we described previously (the selection of nodes, items and edges), also apply here.

If one has access to the network parameters, for example by estimating them from time-series data, the expected state can be estimated by simulating observations from the network and estimating the mean number of active symptoms. Simulating observations also requires knowing the *process* that governs the network dynamics (i.e., how symptoms activate each other). Unfortunately, the typical idiographic network models estimated from time series, such as Vector Autoregressive models (VAR models), are often linear models with corresponding dynamics that will not result in dynamical landscapes with alternative stable states. A model that does allow for non-linear processes (e.g., sudden jumps from a healthy state to a depressed state) and alternative stable states is the Ising model. Currently, network models based on the Ising model can only be estimated from cross-sectional data as this model does not assume time-lagged dependent observations [114]. However, suppose that future research solves this practical issue and one would have an Ising model that approximates a person's system, then one could add the dynamics and estimate the landscape for that person.

For example, Lunansky et al. [71] use the dynamic regime of Glauber dynamics for the Ising model [52] which allows nodes to (de)activate with a probability that depends on the network and state of the other nodes. For some specific networks the most likely state can be solved analytically (e.g., for the Ising model up to 10 nodes, see [40]). When simulating observations from the network with certain dynamics, one can calculate a probability distribution with probabilities for all of the different states as projected by the specified model.[10] Using this probability distribution, one could calculate the mean number of active nodes, which is called the *projected state*, and use this as an estimate of the expected state of the network. That is, whereas the expected state is the mean number of active symptoms under the actual system with its dynamics, the projected state is the mean number of active symptoms under a specified parameterized network with specified dynamics (e.g., a network Ising model estimated from data, to which Glauber dynamics have been added). To the extent that this specified network and dynamics approximate the data-generating system and dynamics, the projected state forms a good measure of the expected state.

4.4. Targeting the Vulnerability of the Landscape

The expected state discussed in the previous section is a property of the dynamical landscape. That is, if we use the ball-and-cup metaphor again, the expected state is not where the ball is actually located, but where it is expected to be located given the landscape.The landscape also depicts how easily the ball moves back to a healthy or depressed state after a perturbation, i.e., after pushing the ball to another place in the landscape. As such, another property of the landscape is the *resilience* of the system, which is defined as the maintenance of healthy psychological functioning despite facing adversity, such as daily stressors or life events [102,63]. A resilient network is characterized by a landscape in which the expected state is healthy and stable, in the sense that perturbations do not easily push a person into a depressed state, and if one does get into a depressed state it is easier to move back into a healthy state. The opposite landscape, in which perturbations easily move a person into a stable depressed state, in which they get stuck despite

[10] This probability distribution of the states can also be used to measure alternative targets such as the most probable state.

positive perturbations (e.g., treatment interventions), could be considered “vulnerability to depression”. This target is visualized in the fourth row of Fig. 2.

Scheffer et al. [101] describe the importance of resilience in the context of eco-system management, explaining that although the focus is often on controlling disturbances on the system, instead it would be better to reduce the risk of unwanted state shifts by focussing on changes that make the system more resilient:

> “Stochastic events that trigger state shifts (such as hurricanes, droughts or disease outbreaks) are usually difficult to predict or control. Therefore, building and maintaining resilience of desired ecosystem states is likely to be the most pragmatic and effective way to manage ecosystems”

Similarly, conceptualizing depression as a stable state that should be avoided might encourage a focus on controlling external risk factors or disturbances that could push someone into a depressed state, whereas targeting ‘vulnerability to depression’ might encourage a focus on changing the network structure such that the person is more resilient.

As such, the vulnerability of the landscape that characterizes a person’s network can be an important target for measurement in psychology, for example, to assess whether a network has structurally become more resilient after some treatment. Instead of only comparing symptom activity before and after treatment, one could compare the landscape’s vulnerability before and after treatment to assess whether the post-treatment landscape shows more resilience against future stressful events.

4.4.1. Considerations for Measuring the Vulnerability of the Landscape

At the moment, we observe several strategies to assess the landscape of networks. The first strategy is to consider how the projected state of the network changes under perturbations by using simulations. Using the ball-and-cup metaphor depicted in Fig. 1, the expected state is the place where the ball is expected to be given the network when there are no perturbations. To investigate how stable that state is, we can analyze how easily the ball moves back to that state when forced to be in a different state due to perturbations. Thus, for a specific network and assumed dynamics, one can simulate the projected state and assess how stable that state is against simulated perturbations.

A method to implement this strategy is proposed by Lunansky et al. [71], who simulate the projected state of an Ising network using Glauber dynamics, and once every so many iterations implement a perturbation by forcing all the symptoms to be active in the case of a healthy projected state (called an *aggravating perturbation*) or forcing all symptoms to be inactive in the case of a depressed projected state (called an *alleviating perturbation*). As such, they assess the network’s landscape by concluding what the network’s projected state is and how stable that state is against perturbations.

Because estimating the vulnerability of the landscape, just like the projected state, requires knowledge of the network parameters, also here all of the challenges discussed for measuring the network structure apply. In addition, it is not evident what dimension best captures resilience/vulnerability. In Lunansky et al.’s (2022) simulations, resilience had two dimensions: the projected state (healthy vs. unhealthy) and the stability of this state (stable vs. unstable). Lunansky et al. [71] used these dimensions to identify four categories (healthy vs unhealthy projected state $\times$ stable vs unstable), in which a network characterized by a resilient landscape shows a projected state that is healthy and stable. Conversely, a network characterized by a vulnerable landscape shows a projected state that is unhealthy and stable.

But, there are other ways to index vulnerability in a network. For example, one could let go of the four categories and the clinical cut-off and instead only take the mean number of active symptoms given the model (i.e, the projected state) and assess how much this mean number changes under different perturbations (e.g., alleviating and aggravating perturbations). Giving an exact measure of how to compare the projected state with perturbations is beyond the scope of this article, but this is to show that there are different possible avenues for measuring vulnerability. The above strategies are currently only implemented in methods for the Ising network model [114], which means that, for now, the limitations to measure individual landscapes hold as discussed before. However, we are hopeful that future research could develop idiographic approaches to these strategies.

In certain cases, the landscape does not need to be assessed from simulations but can be computed analytically. Therefore, the second strategy is to compute a *potential landscape* which measures how stable every possible state of the network is, and as such, shows how the metaphorical ball would roll over this network landscape [30]. A solution to compute the potential landscape of multivariate psychological systems has been suggested by Cui et al. [29], and this solution has recently been implemented in a method (with a corresponding R-package *Isinglandr*) to compute the landscapes of Ising network models [30]. Using this computed landscape, Cui et al. [30] propose a novel metric to assess the resilience of the network by comparing the stability of the healthy state with that of the depressed state. One could determine the cut-off value between the two states quantitatively, using the shape of the landscape (in the case that the landscape shows bistability), or based on clinical significance such as the DSM symptom sum score for receiving a diagnosis for depression. Comparing the local minimum for the healthy and depressed state offers a new way to indicate the resilience of networks.

In addition to simulation and analytical approaches that are based on an assumed model, one can also assess dynamical properties of the system through experimental manipulations or by using generic indicators of vulnerability in observational studies. In experimental approaches to assess the vulnerability of a system, one may measure symptom activation before an experimental intervention and measure the time it takes the system to return to the baseline activation. For example, Mobach et al. [87] applied a cognitive training intervention to individuals with social anxiety disorder, and measured how quickly the individuals returned to baseline levels of anxiety to use this measure as an operationalization of resilience (i.e., benefitting longer from positive disturbances means less vulnerability to the high anxiety state). This can be interpreted as an attempt to measure features of the dynamical landscape of the individuals in question.

In naturalistic studies, where one cannot use controlled experimentation, generic indicators of stability can be used to assess resilience [99,100,102]. These resilience indicators are based on the idea that systems that lack resilience will show particular behavior; for instance, when a vulnerable system is near a critical transition, the system may show increased autocorrelation of the system state, a phenomenon known as critical slowing down [100,102]. Resilience indicators can be computed from time series, as for instance obtained through ESM data in psychology [117]. In several applications in the area of psychopathology, researchers have indeed associated generic indicators of resilience with the risk of impending episodes of psychopathology using empirical data [115,118–120,68]. However, it has also been noted that the assessment of resilience indicators is far from trivial without adequate knowledge of the system, and that further research is necessary to evaluate the merits of these approaches in psychopathology research [31].

4.5. Targeting Stable States

In some of the literature, depression, or mental disorders more generally, are described as dynamical systems characterized by two stable states: a healthy state and a disordered or depressed state (see e.g., [25,91,109,115]. In this conceptualization, one can for example speak of the ‘onset’ and ‘termination’ of depression or a depressive episode, which refers to the shift into and out of the depressed state. One way of creating a dimension on which one could compare people with each other, or before and after treatment, is a dichotomous one: whether someone is currently in a healthy state versus depressed state (in Fig. 1 this amounts to measuring whether the ball is in the one or the other cup). This target is visualized in the fifth row of Fig. 2. Whereas the target 'current state' concerns symptom activation at the moment of measurement, and the targets 'expected state' and 'landscape vulnerability' concern properties of the dynamical landscape, the stable state is a combination: it is the state that someone is *currently* in, but considers whether this is a healthy or depressed stable state, which requires knowledge of the stability of the states and some cut-off of what is considered 'healthy' and 'depressed'. In contrast to the expected state and vulnerability, two people who are characterized by the same network, and thus are characterized by the same landscape, can still differ in their stable state, because one might, at the moment of measurement, be located in the healthy state, while the other is located in the depressed state. For example, in Fig. 2, person p_2 has the same network structure and landscape as person p_3, but because for person p_3 currently 5 symptoms are active (current state) the person is in a depressed stable state, while p_2, for whom currently only 3 symptoms are active, is in a healthy stable state. This could for example be explained by greater external stress on p_3's system.

4.5.1. Considerations for Measuring Stable States

To measure whether someone is in a healthy stable state or in a depressed stable state requires both measuring in what state someone currently is (healthy or depressed), and whether that state is stable. Because a state can be any number ranging from 0 to the total number of symptoms in the network, dividing these states over healthy and depressed states requires some cut-off from which one concludes that the number of symptoms that are active should be considered "depressed". When people show gradual increase or decrease in the number of symptoms, this cut-off is rather arbitrary (one could for example use some clinical diagnostic cut-off such as the DSM; [71]). However, this becomes less arbitrary when people show sudden jumps between stable states of which one is characterized by relatively few active symptoms and one is characterized by relatively many active symptoms.

One possibility is to estimate the landscape and identify stable states from this landscape. The exact meaning of the horizontal and vertical axes of an estimated stability landscape will depend on the specific mathematical procedure used to estimate the landscape. One example of how such a landscape could be estimated is given by Cui et al. [30]. This example of estimating a potential landscape was also discussed before, for measuring depression vulnerability. In this landscape, the horizontal axis represents the possible macrostates (i.e., sum score of active symptoms) and the vertical axis represents the outcome of the potential function, which represents the potential energy in these states, for which it is the case that the system will tend to go to states of lower energy making such states more likely (for details on the exact mathematical procedure used to estimate such a potential landscape we refer the reader to the paper by [30], as we discuss their solution as merely one possibility of estimating a stability landscape). This estimated landscape provides the stability of each possible state, where lower outcomes of the potential function indicate more stability, and as such measures the landscape in which the estimated network is operating. If one would recognize two stable states in this potential landscape, of which one is characterized by low symptom activation and one by high symptom activation, the cut-off between healthy and depressed could be chosen somewhere in between. For example, Cui et al. [30] estimate a network from empirical data on nine depression symptoms (presence/absence), and use this network to compute a potential landscape. The resulting potential landscape shows two 'cups' (stable states), of which one has a low number of symptoms and one has a higher number of symptoms which leads the authors to call one the 'healthy phase' and one the 'depressive phase'. By locating someone's current state (the number of symptoms that are present at the moment of measurement) in this potential landscape, one could evaluate whether the person is located in a healthy stable state, depressed stable state or, although less likely, happens to be in an area that is unstable. As was discussed for the target 'vulnerability', a big challenge for estimating the landscape is that currently the potential landscape method only works for Ising models which can be estimated from cross-sectional data, and so in order to obtain individual landscapes, methods need to be developed that can estimate landscapes from individual networks [30]. In addition, because this target requires measurement of the current state, all of the challenges that hold for measuring the current state, also hold here.

5. Discussion

In the first half of this article, we showed that the assumptions in psychometric models, and as such also the normative criteria that follow from these assumptions, are closely connected with how the measurement target is conceptualized. Because the network framework offers new conceptualizations of psychological constructs, we also expect new psychometric criteria. That is, common normative criteria of traditional psychometric theories, for instance when it comes to the selection of items, or what part of the variance is considered 'error', follow from assumptions that need not hold in the network framework. Because normative criteria for measurement in the network framework will depend on what property is targeted for measurement, we used the second half of this article to explore several possible measurement targets in the network framework.

Using the case of depression as an example of a multidimensional system, we suggested five possible targets: the network structure, current and expected node activity, vulnerability to getting in a depressed state, and whether a person is in a healthy or depressed state. All of the targets share as a big challenge the selection of nodes. In a way, one could say that the focus on selection of items in other psychometric frameworks, shifts to the selection of nodes in the network framework. For the measurement of the nodes one could use existing criteria, such as selecting items that are reliable indicators of the nodes, and in case there is the space in a questionnaire or test, one can use multiple items for each or some of the nodes to improve reliability. The selection of nodes, however, seems a more difficult issue that requires further research into which nodes or combinations of nodes give the best indication of the target property (e.g., the state of a (partly) unobserved network, or dynamical behavior of the system). Such research could then result in the development of criteria for the selection of nodes for a certain target.

Another important conclusion is that this is not an either-or situation; each of the considered targets could be a relevant target of measurement in a different context. The targets also map onto slightly different concepts all related to depression: in the network framework, 'depression' is often used to denote the system or network, current and expected node activity lend themselves to an interpretation of 'depression severity', 'depressive episode' could be understood as being in a depressed stable state, and 'vulnerability to depression' could be understood as the vulnerability of the dynamical landscape.

Distinguishing between these different properties might also be helpful in mapping them onto different ways in which 'depression' is used in the literature. For example, in work on genetic and epigenetic risk factors for depression (e.g., [97]), depression is probably understood as something like vulnerability to depression, as it is unlikely that genetic factors predict whether one is currently depressed (being in a depressed as opposed to healthy state) but rather predict whether one has a high risk of developing depression. Similarly, when investigating the role of childhood trauma in developing depression later in life (e.g., [89]), it is vulnerability to depression that seems to be the most relevant property. In the ball-and-cup metaphor, such predictors of depression can be hypothesized to shape the landscape, rather than affecting where the ball is located at a certain moment in time.

On the other hand, studies that look at recent life events or social support (both supportive and stressful qualities of the social environment) reported before onset of depression (e.g., [93]) as well as studies that look at early warning signals for predicting onset of depression (e.g., [120]), might map better onto the depressed vs healthy stable state as a measurement target. For therapeutic interventions in the network framework, one could distinguish between interventions on nodes, such as sleep medication as a perturbation on the node 'sleeping problems' (trying to push a person into a healthy state) as well as interventions on the network structure, such as trying to break links between symptoms (shaping a more resilient landscape).

We have focused on psychopathology, with depression as an example, because network theory has been most extensively applied to psychopathology. While the explored targets might have equivalents for other disorders (e.g., healthy vs disordered states, and vulnerability to the disorder) they might not have sensible equivalents for other constructs in psychology (e.g., intelligence, personality, or attitudes). The take home message should therefore rather be that possible normative criteria will depend on what it is one wants to measure in the network framework, and that this could involve different properties depending on the substantive context. That is, there are different possible targets for depression, but there might be yet different possible targets for other constructs. It would therefore be interesting to explore, in a similar way, what measurement targets could be relevant in a network perspective on cognitive abilities or personality dimensions.

The measurement targets we considered can all be understood as properties. The first measurement target we considered, the network structure, is a property of a population (in case the relations in the network reflect intraindividual relations) or of a person (in case of individual networks in which the relations reflect within-subject relations). The network structure is a nominal property because the structures that characterize people or populations are not ordered (i.e., individuals with the same network structure make up an equivalence class). So, for example, just like 'birth place' refers to the location a person is born, 'network structure' refers to the network structure that characterizes a person.

Importantly, however, other measurement targets we considered do allow for ordering: people can have more or less symptom activation at the time of measurement (current state), people can be expected to have more or less symptom activation given their landscape (expected state), and people can be more or less vulnerable to a depressed state given their landscape (landscape vulnerability).Whereas the current state and expected state are closer to an interpretation of depression severity, landscape vulnerability can be interpreted as how vulnerable the person is to get into depressive episodes. The stable state as a binary variable does not allow for ordering but rather categorizes people into those who are in a healthy state and those who are in a depressed state. All measurement targets represent properties on which people can vary from each other as well as over time.

The list of possible measurement targets we suggested is by no means an exhaustive list. Also for the example of depression, there can be multiple alternative targets. For example, a target related to that of the stable state which is already being targeted in the literature is how close someone is to transitioning from one state to another (see the literature on early warning signals; [115,118,120]). Similarly, the notion of rigidity has been proposed to capture the degree to which a person is "trapped" in a particular state [105]. Our conclusion is not that there are five measurement targets in the network framework, but rather, first, that existing measurement theories incorporate normative criteria that do not fit well with network conceptualizations of psychological constructs, and second, that there are different dimensions that can be targeted for measurement in the network framework, and all come with their own challenges for measurement.

Measurement of network constructs requires clarity on what the relevant property is to target, so that normative criteria can be developed for this target that prescribe how to best measure it. Given that our survey of the possibilities suggests that contextual features of the substantive context will likely determine which normative criteria are important, we have refrained from articulating a one-size-fits-all psychometric theory, a set of concrete steps for test design, or the suggestion of uniform standards. Although for certain specific models (e.g., the Ising model; [79]) one could imagine that a uniform psychometric approach could be formulated, and it would be certainly useful to pursue such a system, in general we expect that different network constructs (e.g., psychopathology, cognitive ability, personality) and different properties (e.g., the targets discussed for depression) will differ in what measurement theory fits best to measure that target. As such, consideration of the network under investigation and exploring what property one aims to target, will be necessary to determine how this property is best measured.

It is clear that several fields in psychology are undergoing a theoretical shift towards an understanding of constructs in terms of complex networks, and the many novel measurement approaches discussed in this article may well be viewed as a response to that shift. This suggests that we may be witnessing a process of epistemic iteration [20], where measurement practices are changed in response to changing construct theories, although further historical and conceptual research could investigate this process in greater detail. The exploration into possible relevant measurement targets for the example of depression revealed multiple challenges, but we are optimistic and view these challenges and the accompanying open questions as a research agenda for future work, one that will eventually result in one or more measurement theories for the network framework.

CRediT authorship contribution statement

Riet van Bork: Writing – review & editing, Writing – original draft, Visualization. **Gabriela Lunansky:** Writing – review & editing, Writing – original draft. **Denny Borsboom:** Writing – review & editing, Writing – original draft, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work has been supported by NWO Vici Grant 181.029.

References

[1] E.B. Andersen, Sufficient statistics and latent trait models, Psychometrika 42 (1) (1977) 69–81, https://doi.org/10.1007/BF02293746.
[2] A.T. Beck, C.H. Ward, M. Mendelson, J. Mock, J. Erbaugh, An inventory for measuring depression, Arch. Gen. Psychiatry 4 (6) (1961) 561–571, https://doi.org/10.1001/archpsyc.1961.01710120031004.
[3] K.A. Bollen, Structural equations with latent variables, Wiley, New York, 1989.
[4] K.A. Bollen, S. Bauldry, Three Cs in measurement models: causal indicators, composite indicators, and covariates, Psychol. Methods 16 (3) (2011) 265–284, https://doi.org/10.1037/a0024448.
[5] K.A. Bollen, R. Lennox, Conventional wisdom on measurement: a structural equation perspective, Psychol. Bull. 110 (1991) 305–314, https://doi.org/10.1037/0033-2909.110.2.305.
[6] K.A. Bollen, J. Pearl, Eight myths about causality and structural equation models (2013) 301–328, https://doi.org/10.1007/978-94-007-6094-3_15.
[7] T.G. Bond, C.M. Fox, Applying the rasch model: fundamental measurement in the social sciences, Lawrence Erlbaum Associates, Mahwah, NJ, 2001.
[8] D. Borsboom, Measuring the mind: conceptual issues in contemporary psychometrics, Cambridge University Press, 2005.
[9] D. Borsboom, A network theory of mental disorders, World Psychiatry 16 (1) (2017) 5–13, https://doi.org/10.1002/wps.20375.
[10] D. Borsboom, A.O. Cramer, V.D. Schmittmann, S. Epskamp, L.J. Waldorp, The small world of psychopathology, PLoS One 6 (2011) e27407.

[11] D. Borsboom, M.K. Deserno, M. Rhemtulla, S. Epskamp, E.I. Fried, R.J. McNally, L.J. Waldorp, Network analysis of multivariate data in psychological science, Nature Reviews Methods Primers 1 (1) (2021) 58, https://doi.org/10.1038/s43586-021-00055-w.

[12] D. Borsboom, E.I. Fried, S. Epskamp, L.J. Waldorp, C.D. Van Borkulo, H.L. van der Maas, A.O. Cramer, False alarm? A comprehensive reanalysis of "Evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger (2017), https://doi.org/10.1037/abn0000306.

[13] D. Borsboom, G.J. Mellenbergh, J. Van Heerden, The theoretical status of latent variables, Psychol. Rev. 110 (2) (2003) 203–219, https://doi.org/10.1037/0033-295X.110.2.203.

[14] L. Boschloo, C.D. van Borkulo, D. Borsboom, R.A. Schoevers, A prospective study on how symptoms in a network predict the onset of depression, Psychother. Psychosom. 85 (3) (2016) 183–184, https://doi.org/10.1159/000442001.

[15] R.L. Brennan, Generalizability Theory, Springer, New York, 2001.

[16] L.F. Bringmann, C. Albers, C. Bockting, D. Borsboom, E. Ceulemans, A. Cramer, M. Wichers, Psychopathological networks: Theory, methods and practice, Behav. Res. Ther. 149 (2022) 104011, https://doi.org/10.1016/j.brat.2021.104011.

[17] L.F. Bringmann, L.H. Lemmens, M.J. Huibers, D. Borsboom, F.J.P.M. Tuerlinckx, Revealing the dynamic network structure of the beck depression inventory-II, Psychol. Med. 45 (4) (2015) 747–757, https://doi.org/10.1017/S0033291714001809.

[18] L.F. Bringmann, N. Vissers, M. Wichers, N. Geschwind, P. Kuppens, F. Peeters, F. Tuerlinckx, A network approach to psychopathology: new insights into clinical longitudinal data, PLoS One 8 (4) (2013) e60188.

[19] J. Burger, A.-M. Isvoranu, G. Lunansky, J.M.B. Haslbeck, S. Epskamp, R.H. A. Hoekstra, E.I. Fried, D. Borsboom, T.F. Blanken, Reporting standards for psychological network analyses in cross-sectional data, in: Psychological Methods, American Psychological Association, 2023, pp. 806–824, https://doi.org/10.1037/met000047.

[20] H. Chang, Inventing temperature: measurement and scientific progress, Oxford University Press, 2004.

[21] A.P. Christensen, H. Golino, Statistical equivalency of factor and network loadings, PsyArXiv (2020), https://doi.org/10.31234/osf.io/xakez.

[22] A.P. Christensen, H. Golino, P.J. Silvia, A psychometric network perspective on the validity and validation of personality trait questionnaires, Eur. J. Pers. 34 (6) (2020) 1095–1108, https://doi.org/10.1002/per.2265.

[23] D.M. Condon, The SAPA personality inventory: An empirically-derived, hierarchically-organized self-report personality assessment model, PsyArXiv (2018), https://doi.org/10.31234/osf.io/sc4p9.

[24] D.R. Cox, N. Wermuth, Linear dependencies represented by chain graphs, Stat. Sci. 8 (1993) 204–218, https://doi.org/10.1214/ss/1177010887.

[25] A.O.J. Cramer, C.D. Van Borkulo, E.J. Giltay, H.L.J. Van Der Maas, K.S. Kendler, M. Scheffer, D. Borsboom, Major depression as a complex dynamic system, PLoS One (2016), https://doi.org/10.1371/journal.pone.0167490.

[26] A.O.J. Cramer, S. Van der Sluis, A. Noordhof, M. Wichers, N. Geschwind, S. H. Aggen, D. Borsboom, Dimensions of normal personality as networks in search of equilibrium: you can't like parties if you don't like people, Eur. J. Pers. 26 (4) (2012) 414–431, https://doi.org/10.1002/per.1866.

[27] A.O.J. Cramer, L.J. Waldorp, H. van der Maas, D. Borsboom, Authors' response. Complex realities require complex theories: refining and extending the network approach to mental disorders, Behav. Brain Sci. 33 (2010) 178–193, https://doi.org/10.1017/S0140525X10000920.

[28] L.J. Cronbach, G.C. Gleser, H. Nanda, N. Rajaratnam, The dependability of behavioral measurements: theory of generalizability scores and profiles, Wiley, 1972.

[29] J. Cui, A. Lichtwarck-Aschoff, M. Olthof, T. Li, F. Hasselman, From metaphor to computation: constructing the potential landscape for multivariate psychological formal models, Multivar. Behav. Res. 1–19 (2022), https://doi.org/10.1080/00273171.2022.2119927.

[30] J. Cui*, G. Lunansky*, A. Lichtwark-Aschoff, N.B. Mendoza, F. Hasselman, Quantifying the stability of psychological networks: a potential landscape approach, PsyArxiv (2023), https://doi.org/10.31234/osf.io/nd8zc.

[31] F. Dablander, A. Pichler, A. Cika, A. Bacilieri, Anticipating critical transitions in psychological systems using early warning signals: theoretical and practical considerations, Psychol. Methods (2022), https://doi.org/10.1037/met0000450 (Advance online publication. doi: 10.1037/met0000450).

[32] J. Dalege, D. Borsboom, F. van Harreveld, H. van den Berg, M. Conner, H.L.J. van der Maas, Toward a formalized account of attitudes: the causal attitude network (CAN) model, Psychol. Rev. 123 (1) (2016) 2–22, https://doi.org/10.1037/a0039802.

[33] N.S. De Boer, L.C. de Bruin, J.J. Geurts, G. Glas, The network theory of psychiatric disorders: a critical assessment of the inclusion of environmental factors, Front. Psychol. 12 (2021) 623970, https://doi.org/10.3389/fpsyg.2021.623970.

[34] P. De Boeck, Random item IRT models, Psychometrika 73 (2008) 533–559, https://doi.org/10.1007/s11336-008-9092-x.

[35] M.K. Deserno, D. Borsboom, S. Begeer, R. van Bork, M. Hinne, H.M. Geurts, Highways to happiness for autistic adults? perceived causal relations among clinicians, PLoS One 15 (12) (2020) e0243298.

[36] J. De Ron, D.J. Robinaugh, E.I. Fried, P. Pedrelli, F.A. Jain, D. Mischoulon, S. Epskamp, Quantifying and addressing the impact of measurement error in network models, Behav. Res. Ther. 157 (2022) 104163, https://doi.org/10.1016/j.brat.2022.104163.

[37] C.V. Dolan, D. Borsboom, Interpretational issues with the bifactor model: a commentary on 'Defining the p-Factor: an empirical test of five leading theories' by southward, cheavens, and coccaro, Psychol. Med. 53 (2023) 2744–2747, https://doi.org/10.1017/S0033291723000533.

[38] J.R. Edwards, R.P. Bagozzi, On the nature and direction of relationships between constructs and measures, Psychol. Methods 5 (2) (2000) 155–174, https://doi.org/10.1037/1082-989X.5.2.155.

[39] J.L. Ellis, B.W. Junker, Tail-measurability in monotone latent variable models, Psychometrika 62 (4) (1997) 495–523, https://doi.org/10.1007/bf02294640.

[40] S. Epskamp, IsingSampler: Sampling Methods and Distribution Functions for the Ising Model (R package version 0.2.1.) (2020), https://cran.r-project.org/web/packages/IsingSampler/IsingSampler.pdf.

[41] S. Epskamp, D. Borsboom, E.I. Fried, Estimating psychological networks and their accuracy: a tutorial paper, Behav. Res. Methods 50 (1) (2018) 195–212, https://doi.org/10.3758/s13428-017-0862-1.

[42] S. Epskamp, E.I. Fried, A tutorial on regularized partial correlation networks, Psychol. Methods 23 (4) (2018) 617–634, https://doi.org/10.1037/met0000167.

[43] S. Epskamp, A.M. Isvoranu, M.W.L. Cheung, Meta-analytic Gaussian network aggregation, Psychometrika 87 (2022) 12–46, https://doi.org/10.1007/s11336-021-09764-3.

[44] S. Epskamp, G. Maris, L.J. Waldorp, D. Borsboom, Network psychometrics, in: P. Irwing, T. Booth, D.J. Hughes (Eds.), The Wiley Handbook of Psychometric Testing, 2 Volume Set: A Multidisciplinary Reference on Survey, Scale and Test Development, Wiley, New York, NY, 2018, pp. 953–986, https://doi.org/10.1002/9781118489772.ch30.

[45] S. Epskamp, M. Rhemtulla, D. Borsboom, Generalized network psychometrics: combining network and latent variable models, Psychometrika 82 (2017) 904–927, https://doi.org/10.1007/s11336-017-9557-x.

[46] A.J. Fisher, J.D. Medaglia, B.F. Jeronimus, Lack of group-to-individual generalizability is a threat to human subjects research, Proc. National Academy Sci. United States of America (2018), https://doi.org/10.1073/pnas.1711978115.

[47] A. Finnemann, D. Borsboom, S. Epskamp, H.L.J. van der Maas, The theoretical and statistical Ising model: a practical guide in R, Psych 3 (4) (2021) 594–618, https://doi.org/10.3390/psych3040039.

[48] M.K. Forbes, A.G. Wright, K.E. Markon, R.F. Krueger, Evidence that psychopathology symptom networks have limited replicability, J. Abnorm. Psychol. 126 (7) (2017) 969, https://doi.org/10.1037/abn0000276.

[49] R. Foygel, M. Drton, Extended Bayesian information criteria for Gaussian graphical models, Adv. Neural Inf. Proces. Syst. 23 (2010) 2020–2028.

[50] P.A. Frewen, S.L. Allen, R.A. Lanius, R.W. Neufeld, Perceived causal relations: novel methodology for assessing client attributions about causal associations between variables including symptoms and functional impairment, Assessment 19 (2012) 480–493, https://doi.org/10.1177/1073191111418297.

[51] K.M. Gates, P.C.M. Molenaar, Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples, Neuroimage 63 (1) (2012) 310–319, https://doi.org/10.1016/j.neuroimage.2012.06.026.

[52] R.J. Glauber, Time-dependent statistics of the ising model, J. Math. Phys. 4 (2) (1963) 294–307, https://doi.org/10.1063/1.1703954.

[53] L. Guttman, A basis for scaling qualitative data, Am. Sociol. Rev. 9 (1944) 139–150, https://doi.org/10.2307/2086306.

[54] L. Guttman, The basis for scalogram analysis, in: S.A. Stouffer, L. Guttman, E. A. Suchman, P.F. Lazarsfeld, S.A. Star, J.A. Claussen (Eds.), Measurement and prediction, Princeton University Press, Princeton NJ, 1950, pp. 60–90.

[55] J.M.B. Haslbeck, L.J. Waldorp, How well do network models predict observations? on the importance of predictability in network models, Behav. Res. Methods 50 (2018) 853–861, https://doi.org/10.3758/s13428-017-0910-x.

[56] R.H.A. Hoekstra, S. Epskamp, D. Borsboom, Heterogeneity in Individual Network Analysis: Reality or Illusion?, Multivariate Behav. Res. (2022), Advance online publication, https://doi.org/10.1080/00273171.2022.2128020.

[57] T.R. Insel, B.N. Cuthbert, Brain disorders? precisely, Science 348 (2015) 499–500, https://doi.org/10.1126/science.aab2358.

[58] E. Ising, Beitrag zur theorie des ferromagnetismus, Zeitschrift für Physik A Hadrons and Nuclei 31 (1) (1925) 253–258, https://doi.org/10.1007/bf02980577.

[59] A.M. Isvoranu, S. Epskamp, Which estimation method to choose in network psychometrics? deriving guidelines for applied researchers, Psychol. Methods 28 (4) (2023) 925–946, https://doi.org/10.1037/met0000439.

[60] K.G. Jonas, K.E. Markon, A descriptivist approach to trait conceptualization and inference, Psychol. Rev. 123 (1) (2016) 90–96, https://doi.org/10.1037/a0039542.

[61] P.J. Jones, A. Heeren, R.J. McNally, Commentary: a network theory of mental disorders, Front. Psychol. 8 (2017), https://doi.org/10.3389/fpsyg.2017.01305.

[62] M. Kac, Mathematical mechanisms of phase transitions, in: M. Chrétien, E. Gross, S. Deser (Eds.), Statistical physics: Phase transitions and superfluidity, vol. 1, Brandeis university summer institute in theoretical physics, Gordon and Breach Science Publishers, New York, 1968, pp. 241–305.

[63] R. Kalisch, D.G. Baker, U. Basten, M.P. Boks, G.A. Bonanno, E. Brummelman, B. Kleim, The resilience framework as a strategy to combat stress-related disorders, Nat. Hum. Behav. 1 (11) (2017) 784–790, https://doi.org/10.1038/s41562-017-0200-8.

[64] K.-J. Kan, H. de Jonge, H.L.J. van der Maas, S.Z. Levine, S. Epskamp, How to compare psychometric factor and network models, J. Intelligence 8 (4) (2020) 35, https://doi.org/10.3390/jintelligence8040035.

[65] L. Klintwall, M. Bellander, M. Cervin, Perceived causal problem networks: reliability, central problems, and clinical utility for depression, Assessment 30 (2023) 73–83, https://doi.org/10.1177/10731911211039281.

[66] D.H. Krantz, R.D. Luce, P. Suppes, A. Tversky, Foundations of measurement, I, Academic Press, New York, 1971.
[67] R.F. Krueger, The structure of common mental disorders, Arch. Gen. Psychiatry 56 (10) (1999) 921–926, https://doi.org/10.1001/archpsyc.56.10.921.
[68] Y.K. Kunkels, H. Riese, S.E. Knapen, R.F. Riemersma-van der Lek, S.V. George, A. M. van Roon, R.A. Schoevers, M. Wichers, Efficacy of early warning signals and spectral periodicity for predicting transitions in bipolar patients: an actigraphy study, Transl. Psychiatry 11 (1) (2021) 350, https://doi.org/10.1038/s41398-021-01465-w.
[69] C.A. Levinson, C. Cusack, M.L. Brown, A.R. Smith, A network approach can improve eating disorder conceptualization and treatment, Nat. Rev. Psyc. 1 (7) (2022) 419–430, https://doi.org/10.1038/s44159-022-00062-y.
[70] F.M. Lord, M.R. Novick, Statistical theories of mental test scores, Addison Wesley, 1968.
[71] G. Lunansky, C. van Borkulo, T. Blanken, A.O.J. Cramer, D. Borsboom, Bouncing back from life's perturbations: formalizing psychological resilience from a complex systems perspective, PsyArxiv (2022), https://doi.org/10.31234/osf.io/ftx4j.
[72] L. Luo, Z.F. Fisher, C. Arizmendi, P. Molenaar, A. Beltz, K.M. Gates, Estimating both directed and undirected contemporaneous relations in time series data using hybrid-group iterative multiple model estimation, Psychol. Methods 28 (2023) 189–206, https://doi.org/10.1037/met0000485.
[73] L. Mari, Toward a harmonized treatment of nominal properties in metrology, Metrologia 54 (5) (2017) 784, https://doi.org/10.1088/1681-7575/aa8521.
[74] L. Mari, C. Narduzzi, G. Nordin, S. Trapmann, Foundations of uncertainty in evaluation of nominal properties, Measurement 152 (2020) 107397, https://doi.org/10.1016/j.measurement.2019.107397.
[75] L. Mari, M. Wilson, A. Maul, Measurement across the sciences (2nd Ed.). Springer Series in Measurement, Science and Technology, 2023, https://doi.org/10.1007/978-3-031-22448-5.
[76] K.A. Markus, Three conceptual impediments to developing scale theory for formative scales, Methodology 14 (2018), https://doi.org/10.1027/1614-2241/a000154.
[77] K.A. Markus, D. Borsboom, Frontiers of test validity theory: measurement, causation, and meaning, Routledge, 2013.
[78] M. Marsman, M. Rhemtulla, Guest editors' introduction to the special issue "Network psychometrics in action": Methodological innovations inspired by empirical problems, Psychometrika 87 (1) (2022) 1–11, https://doi.org/10.1007/s11336-022-09861-x.
[79] M. Marsman, D. Borsboom, J. Kruis, S. Epskamp, R. van Bork, L.J. Waldorp, H.L. J. van der Maas, G. Maris, An introduction to network psychometrics: relating Ising network models to item response theory models, Multivar. Behav. Res. 53 (2018) 15–35, https://doi.org/10.1080/00273171.2017.1379379.
[80] M. Marsman, L. Waldorp, D. Borsboom, Towards an encompassing theory of network models: reply to Brusco, Steinley, Hoffman, Davis-Stober, and Wasserman (2019), Psychol. Methods (2022), https://doi.org/10.1037/met0000373.
[81] R.J. McNally, P. Mair, B.L. Mugno, B.C. Riemann, Co-morbid obsessive–compulsive disorder and depression: a Bayesian network approach, Psychol. Med. 47 (7) (2017) 1204–1214, https://doi.org/10.1017/S0033291716003287.
[82] N. Meinshausen, P. Bühlmann, High-dimensional graphs and variable selection with the lasso, The Annals of Statistics 34 (2006) 1436–1462, https://doi.org/10.1214/009053606000000281.
[83] G.J. Mellenbergh, Item bias and item response theory, Int. J. Educ. Res. 13 (2) (1989) 127–143, https://doi.org/10.1016/0883-0355(89)90002-5.
[84] G.J. Mellenbergh, Generalized linear item response theory, Psychol. Bull. 115 (1994) 300–307, https://doi.org/10.1037/0033-2909.115.2.300.
[85] W. Meredith, Measurement invariance, factor analysis and factorial invariance, Psychometrika 58 (1993) 525–543, https://doi.org/10.1007/BF02294825.
[86] J. Michell, Quantitative science and the definition of measurement in psychology, Br. J. Psychol. 88 (3) (1997) 355–383, https://doi.org/10.1111/j.2044-8295.1997.tb02641.x.
[87] L. Mobach, R. van Loenen, E.A. Dam, D. Borsboom, R.W. Wiers, E. Salemink, Return to baseline after an interpretation training as a dynamic predictor for treatment response in social anxiety disorder, Cogn. Ther. Res. 47 (2023) 431–438, https://doi.org/10.1007/s10608-022-10346-5.
[88] P.C. Molenaar, A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever, Measurement 2 (4) (2004) 201–218, https://doi.org/10.1207/s15366359mea0204_1.
[89] V. Moskvina, A. Farmer, V. Swainson, J. O'Leary, C. Gunasinghe, M. Owen, A. Korszun, Interrelationship of childhood trauma, neuroticism, and depressive phenotype, Depress. Anxiety 24 (3) (2007) 163–168, https://doi.org/10.1002/da.20216.
[90] Z.P. Neal, J.W. Neal, Out of bounds? the boundary specification problem for centrality in psychological networks, *Psychol. Methods* Advance online publication. (2021), https://doi.org/10.1037/met0000426.
[91] B. Nelson, P.D. McGorry, M. Wichers, J.T. Wigman, J.A. Hartmann, Moving from static to dynamic models of the onset of mental disorder: a review, JAMA Psychiat. 74 (5) (2017) 528–534.
[92] P.E. Newton, S. Shaw, Validity in educational and psychological assessment, Sage, London, 2014.
[93] E.S. Paykel, Life events, social support and depression, Acta Psychiatr. Scand. 89 (1994) 50–58, https://doi.org/10.1111/j.1600-0447.1994.tb05803.x.
[94] G. Rasch, Probabilistic models for some intelligence and attainment tests, Paedagogiske Institut, 1960.
[95] M. Rhemtulla, R. van Bork, D. Borsboom, Worse than measurement error: consequences of inappropriate latent variable measurement models, Psychol. Methods 25 (1) (2020) 30–45, https://doi.org/10.1037/met0000220.
[96] D.J. Robinaugh, R.H. Hoekstra, E.R. Toner, D. Borsboom, The network approach to psychopathology: a review of the literature 2008–2018 and an agenda for future research, Psychol. Med. 50 (3) (2020) 353–366, https://doi.org/10.1017/S0033291719003404.
[97] N.S. Roetker, J.A. Yonker, C. Lee, V. Chang, J.J. Basson, C.L. Roan, C.S. Atwood, Multigene interactions and the prediction of depression in the Wisconsin Longitudinal Study, BMJ Open 2 (4) (2012) e000944.
[98] M. Scheffer, Multiplicity of stable states in freshwater systems, in: R.D. Gulati, E. H.R.R. Lammens, M.L. Meijer, E. van Donk (Eds.), Biomanipulation Tool for Water Management. Developments in Hydrobiology, Springer, Dordrecht, 1990, https://doi.org/10.1007/978-94-017-0924-8_42.
[99] M. Scheffer, Critical Transitions in Nature and Society, Princeton University Press, 2009.
[100] M. Scheffer, J. Bascompte, W.A. Brock, V. Brovkin, S.R. Carpenter, V. Dakos, G. Sugihara, Early-warning signals for critical transitions, Nature 461 (7260) (2009) 53–59, https://doi.org/10.1038/nature08227.
[101] M. Scheffer, S. Carpenter, J.A. Foley, C. Folke, B. Walker, Catastrophic shifts in ecosystems, Nature 413 (6856) (2001) 591–596, https://doi.org/10.1038/35098000.
[102] M. Scheffer, J.E. Bolhuis, D. Borsboom, S. Gijzel, D. Goulson, J. Kammenga, B. Kemp, I. Van de Leemput, S. Levin, C. Mary Martin, R. Melis, E. Van Nes, M. Romero, M. Olde Rikkert, Quantifying resilience of humans and other animals, Proc. Natl. Acad. Sci. (2018), https://doi.org/10.1073/pnas.1810630115.
[103] M. Schmitt, R. Steyer, A latent state-trait model (not only) for social desirability, Pers. Individ. Differ. 14 (4) (1993) 519–529, https://doi.org/10.1016/0191-8869(93)90144-r.
[104] L.A. Selby-Bigge (Ed.), Enquiries Concerning Human Understanding and Concerning the Principles of Morals, Oxford University Press, Oxford, 1975.
[105] M.N. Servaas, R.A. Schoevers, L.F. Bringmann, M.-J. van Tol, H. Riese, Trapped: rigidity in psychiatric disorders, Lancet Psychiatry 8 (12) (2021) 1022–1024, https://doi.org/10.1016/S2215-0366(21)00353-9.
[106] K. Sijtsma, On the use, the misuse, and the very limited usefulness of Cronbach's alpha, Psychometrika 74 (2009) 107–120, https://doi.org/10.1007/s11336-008-9101-0.
[107] C. Spearman, "General intelligence", objectively determined and measured, Am. J. Psychol. 15 (2) (1904) 201–292, https://doi.org/10.2307/1412107.
[108] W.F. Stout, A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation, Psychometrika 55 (1990) 293–325, https://doi.org/10.1007/BF02295289.
[109] T.J. Strauman, Modeling the onset of a depressive episode: a self-regulation perspective, Curr. Opin. Psychol. 41 (2021) 100–106, https://doi.org/10.1016/j.copsyc.2021.04.003.
[110] D. Thielemann, M.-A. Sengewald, G. Kappler, R. Steyer, A probit latent state irt model with latent item-effect variables, Eur. J. Psychol. Assess. 33 (4) (2017) 271–284, https://doi.org/10.1027/1015-5759/a000417.
[111] P. Tio, S. Epskamp, A. Noordhof, D. Borsboom, Mapping the manuals of madness: comparing the ICD-10 and DSM-IV-TR using a network approach, Int. J. Methods Psychiatr. Res. 25 (2016) 267–276, https://doi.org/10.1002/mpr.1503.
[112] R. Van Bork, M. Rhemtulla, K. Sijtsma, D. Borsboom, A causal theory of error scores, Psychol. Methods (2022), Advance online publication, https://doi.org/10.1037/met0000521.
[113] R. Van Bork, L.D. Wijsen, M. Rhemtulla, Toward a causal interpretation of the common factor model, Disputatio 9 (47) (2017) 581–601, https://doi.org/10.1515/disp-2017-0019.
[114] C.D. Van Borkulo, D. Borsboom, S. Epskamp, T.F. Blanken, L. Boschloo, R. A. Schoevers, L.J. Waldorp, A new method for constructing networks from binary data, Sci. Rep. 4 (2014) 5918, https://doi.org/10.1038/srep05918.
[115] I.A. Van de Leemput, M. Wichers, A.O.J. Cramer, D. Borsboom, F. Tuerlinckx, P. Kuppens, E.H. van Nes, W. Viechtbauer, E.J. Giltay, S.H. Aggen, C. Derom, N. Jacobs, K.S. Kendler, H.L.J. van der Maas, M.C. Neale, F. Peeters, E. Thiery, P. Zachar, M. Scheffer, Critical slowing down as early warning for the onset and termination of depression, National Academy of Sciences. Proceedings 111 (1) (2014) 87–92, https://doi.org/10.1073/pnas.1312114110.
[116] H.L.J. Van der Maas, C.V. Dolan, R.P. Grasman, J.M. Wicherts, H.M. Huizenga, M. E. Raijmakers, A dynamical model of general intelligence: the positive manifold of intelligence by mutualism, Psychol. Rev. 113 (4) (2006) 842–861, https://doi.org/10.1037/0033-295X.113.4.842.
[117] M. Wichers, The dynamic nature of depression: a new micro-level perspective of mental disorder that meets current challenges, Psychol. Med. 44 (2014) 1349–1360, https://doi.org/10.1017/s0033291713001979.
[118] M. Wichers, P.C. Groot, Psychosystems, ESM Group, EWS Group, Critical Slowing Down as a Personalized Early Warning Signal for Depression, Psychother. Psychosom. 85 (2016) 114–116, https://doi.org/10.1159/000441458.
[119] M. Wichers, M. Schreuder, R. Goekoop, R. Groen, Can we predict the direction of sudden shifts in symptoms? transdiagnostic implications from a complex systems perspective on psychopathology, Psychol. Med. 49 (3) (2019) 380–387, https://doi.org/10.1017/S0033291718002064.
[120] M. Wichers, A.C. Smit, E. Snippe, Early warning signals based on momentary affect dynamics can expose nearby transitions in depression: a confirmatory

single subject time-series study, Journal for Person-Oriented Research 6 (1) (2020) 1–15, https://doi.org/10.17505/jpor.2020.22042.

[121] L.D. Wijsen, D. Borsboom, A. Alexandrova, Values in psychometrics, Perspect. Psychol. Sci. 17 (3) (2022) 788–804, https://doi.org/10.1177/17456916211014183.

[122] A.K. Wittenborn, H. Rahmandad, J. Rick, N. Hosseinichimeh, Depression as a systemic syndrome: mapping the feedback loops of major depressive disorder, Psychol. Med. 46 (3) (2016) 551–562, https://doi.org/10.1017/S0033291715002044.