# A Psychometric Network Perspective on the Measurement and Assessment of Personality Traits

**Preprint** · March 2019

DOI: 10.31234/osf.io/ktejp



**3 authors:**

Alexander P Christensen
University of North Carolina at Greensboro


Hudson Golino
University of Virginia


Paul J Silvia
University of North Carolina at Greensboro

# A Psychometric Network Perspective on the Measurement and Assessment of Personality Traits

Alexander P. Christensen[1], Hudson F. Golino[2], & Paul J. Silvia[1]

[1] University of North Carolina at Greensboro

[2] University of Virginia

Author Note

Correspondence concerning this article should be addressed to Alexander P. Christensen, Department of Psychology, P.O. Box 26170, University of North Carolina at Greensboro, Greensboro, NC, 27402-6170, USA. E-mail: apchrist@uncg.edu

Abstract

Nearly all personality measurement and assessment has been shaped by traditional psychometric methods such as classical test theory. Despite advances in personality theory and psychometrics, most personality assessment instruments were developed using traditional psychometric concepts and measures. One consequence of these traditional practices is that they present important limitations for addressing the complexity of personality constructs, limiting their application to modern personality theories. In recent years, a new field of network psychometrics has emerged, which embraces the complexity and promises to change the way researchers think about psychological constructs. With this new tool comes the opportunity to refine and advance measurement and assessment practices, particularly in personality research. In this paper, we review modern personality theories and discuss how psychometric network models can inform their conjectures. After, we review modern trends in personality measurement and show how psychometric networks can further advance these movements. Then, we put forward a psychometric network framework for developing personality assessment instruments. Within this framework, we connect network measures to traditional psychometric measures and provide guidelines for how psychometric networks can be used to develop personality scales. Several future directions emphasizing the development and improvement of this framework are provided.

*Keywords:* psychometric networks, personality, measurement, assessment, scale development

# A Psychometric Network Perspective on the Measurement and Assessment of Personality Traits

Psychometric modeling and personality theory have made tremendous progress over the last decade. Assessment instruments (e.g., self-report questionnaires) of personality have not. Most personality questionnaires used today are based on the notions of classical test theory (CTT), despite the long-standing presence of more sophisticated models such as item response theory (IRT; Chernyshenko, Stark, Drasgow, & Roberts, 2007; Embretson, 2004; Lord & Novick, 1968; Simms, 2008). One reason personality researchers have not adopted more modern psychometric practices is that many still adhere to dominant philosophical views of measurement from nearly 60 years ago (Borsboom, 2006). These more traditional perspectives lack the complexity that modern personality theories require. The emerging field of *network psychometrics* embraces this complexity and promises to change the way researchers think about measurement and assessment (Bringmann & Eronen, 2018; Schmittmann et al., 2013).

Psychometric network models, the focus of this paper, are a relatively recent addition to network models and the field of psychometrics more broadly (Epskamp, Maris, Waldrop, & Borsboom, 2018b). Conceptually, network models are simple: nodes (e.g., circles) represent variables (e.g., self-report items), and edges (e.g., lines) represent associations (e.g., correlations) between nodes. This broad definition allows network models to encompass many statistical models already used in psychology (Bringmann & Eronen, 2018; Epskamp, Rhemtulla, & Borsboom, 2017). Structural equation modeling (SEM) and path analysis, for example, are special cases of network models where arrows between (latent) variables are specified (Molenaar, 2010; Molenaar, van Rijn, & Hamaker, 2007).

The theoretical perspective of psychometric network models is often referred to as the *network approach* (Borsboom, 2008, 2017). The network approach suggests that psychological attributes (e.g., personality traits) are complex and dynamic systems of causally coupled behaviors that mutually reinforce one another (Cramer et al., 2012a;

Schmittmann et al., 2013). In this way, personality traits resemble an emergent property of interacting observable behavioral components—that is, traits are not any single component of the system but rather a feature of the system as a whole (Baumert et al., 2017). From this perspective, personality traits are meaningful patterns of covariation at the population-level that arise from the mutualistic interactions between personality characteristics (e.g., facets) and nuances (e.g., items; Cramer et al., 2012a). Thus, traits emerge because some characteristics and processes tend to covary more than others, directly reinforcing one another (Mõttus & Allerhand, 2017).

This perspective is often juxtaposed with latent variable models, suggesting that traits arise from, rather than cause, their constituent behaviors (Edwards & Bagozzi, 2000; Schmittmann et al., 2013). Many researchers, however, are skeptical of whether this contrast is necessary, with some advocating that network and latent variable models are more alike than different (Bringmann & Eronen, 2018; Molenaar, 2010; Molenaar et al., 2007). Indeed, some researchers have shown that certain network models can be made equivalent to certain latent variable models (Epskamp et al., 2018b; Marsman et al., 2018) and can have similar interpretations (Golino & Epskamp, 2017; Kruis & Maris, 2016). Others have provided evidence that certain network measures are roughly equivalent to latent variable measures (Hallquist, Wright, & Molenaar, 2019). This raises the question of whether network models offer anything more than a novel interpretation of personality traits.

The intent of this paper is to demonstrate that network models do offer more than a novel interpretation and that they have the potential to inform personality theory and measurement beyond traditional psychometric approaches. As a consequence, we argue that this substantiates the need for the development of new assessment instruments based on this psychometric perspective. We begin by identifying how network models provide substantive advantages to personality theory and measurement that are not afforded by the other approaches. Afterwards, we present a framework for scale development based on the perspective provided by the network approach. This framework will focus on self-report

questionnaires, which are the most commonly used assessment instrument in personality research. Finally, we discuss future directions for the development of this framework.

## 1. Modern Theories of Personality

The goal of this section is to provide a review of modern personality theories and how they explain the interindividual structure of personality traits—that is, the organization of population-level covariation of broad interindividual personality differences (Baumert et al., 2017). This is not an exhaustive review of personality theories but rather a selection of exemplars. It is also neither a critique nor endorsement for any one theory. Instead, the aim is to provide a background of modern personality theories to demonstrate how psychometric network models can inform their conjectures. Although not all theories presented here require trait explanations, most of the literature assesses personality through the lens of traits; therefore, theories are discussed from this view. Finally, the emphasis on the explanation of trait structure is because although between-person models do not necessarily imply within-person models, between-person models describe taxonomic classifications that can guide and inform within-person measures (Borkenau & Ostendorf, 1998; Borsboom, Mellenbergh, & van Heerden, 2003; Cervone, 2005).

### 1.1. Conventional Trait Theories

Many modern trait theories trace their roots back to the five-factor model (FFM), with the most conventional theories branching directly from the FFM. Conventional theories typically explain traits through a structural lens: traits are the common cause of covariation between characteristics. From this view, traits do not merely describe covariation of behaviors but rather explain why some behaviors covary more than others.

One conventional trait theory comes from McCrae and Costa, who championed the FFM and proposed the Five-Factor Theory (FFT; McCrae & Costa, 1996) as a definition and explanation of the personality system (McCrae & Costa, 2008). The basic tendencies component of the FFT is an explicit explanation of traits and is defined by four postulates

including structure. Traits in the FFT are "organized hierarchically from narrow and specific to broad and general, [with trait domains constituting] the highest level of the hierarchy" (McCrae & Costa, 2008, p. 165). In the FFT, traits are posited as biologically-based properties that are endogenous psychological entities, which generalize to all people such that a person behaves in such a way because they possess some quantity of a trait that causes their behaviors. An alternative perspective is that traits are descriptive phenotypic constructs that do not necessarily cause behaviors; however, this account does not attempt to infer the causal mechanisms of traits (Goldberg, 1993). In general, conventional theories suggest that traits are the overarching cause of personality characteristics and any variability in their expressions is considered error variance.

This theoretical interpretation is conceptually provided by reflective latent variables models, which suggest that a personality trait causes the responses to items (Edwards & Bagozzi, 2000). These models hold the assumption of local independence—that is, after conditioning on the latent variable (i.e., common covariance), the observed variables are statistically independent (i.e., they are no longer correlated). One consequence of these models is that they imply a simple structure of personality (i.e., items and facets are uniquely assigned to higher-order factors; Thurstone, 1947).

In practice, this simple structure has led to unintended issues for the FFM. Many researchers find this structure too simple (Herrmann & Pfister, 2013), some researchers suggest that the latent variable methodology (e.g., confirmatory factor analysis; CFA), rather than the structure, is the issue (McCrae, Zonderman, Costa, Bond, & Paunonen, 1996), while others find this structure to be incompatible with the intraindividual structure of personality (Mischel & Shoda, 1995). Indeed, one problematic assumption occurs when attempting to generalize latent variables to intraindividual processes: Latent variable models assume that the measurement model is the same within and between subjects (Borsboom et al., 2003; Ellis & van den Wollenberg, 1993; Hamaker, Nesselroade, & Molenaar, 2007). To fit within conventional theory frameworks, researchers are forced to invoke traits as an explanation of

intraindividual processes by taking on the assumption that latent theoretical entities cause a person's responses to measures (whether or not they agree with this assumption).

One approach to relax the constraint of a simple structure model is through exploratory Structural Equation Modeling (ESEM; Asparouhov & Muthén, 2009). ESEM is less restrictive than CFA because it allows facets to cross-load onto several personality factors (Herrmann & Pfister, 2013). Indeed, Costa and McCrae (1992) acknowledged that these cross-loadings may be "appropriate and meaningful" (p. 45). Although this approach resolves the simple structure issue, it continues to suggest that interindividual structure generalizes to intraindividual processes.

Psychometric network models offer an alternative modeling solution that potentially mitigates both issues. First, complex latent variable structures can be obtained by integrating latent variables into network models via Latent Network Modeling (LNM; Epskamp et al., 2017). In these models, latent variables are estimated (e.g., facets of a trait) and used as nodes in a network model. The complex structure arises from the associations occuring between the facets of the traits, which can be identified by clusters—sets of connected nodes—in the network (Golino & Epskamp, 2017). Second, psychometric network models offer an integrative framework for connecting within-person processes to between-person structures (Baumert et al., 2017; Beck & Jackson, 2017).

With more recent advancements in network psychometrics, these opposing theoretical interpretations can be tested statistically. Kan, van der Maas, and Levine (2019), for example, extended traditional indices of model fit to psychometric network models, allowing their fit to be compared to CFA models (Kan et al., 2019). In Kan and colleagues' study, they found support for a mutualism model of intelligence (i.e., the network approach; van der Maas et al., 2006) over the common cause model (i.e., latent variable models). Similar approaches could be applied to personality scales to determine whether a common cause model proposed by FFT fits better than models proposed by mutualism theories (e.g., Cramer et al., 2012a). Already, there is some evidence to suggest that a more complex

structure fits the structure of personality traits better than a simple structure (Herrmann & Pfister, 2013).

Juxtaposing latent variable and psychometric network models, however, may not be necessary. Mõttus and Allerhand (2017) provide a potential solution to resolve the theoretical differences with a hypothetical conceptual framework. Their framework suggests that person vectors (i.e., people) tend to settle into particular positions of a multidimensional personality feature space (i.e., population structure of personality). In this feature space, forces (e.g., environmental niches, other people, time-invariant genetic influences) "pull" people towards different areas of the feature space (Mõttus, Allerhand, & Johnson, in pressa). Certain environments, for example, may pull all personality characteristics in a certain direction or they may pull only some. Crucially, latent factors may also represent forces that pull some characteristics to certain areas of the feature space (DeYoung, 2015). Network connections between personality features may then represent a bridge between a person vector's current state and the forces acting on them. In either case, conventional theories can be accommodated within the network approach by compromising traits' causal interpretation (e.g., traits may be only one cause among many).

### 1.2. Socio-cognitive Theories

Another explanation for the covariation of traits comes from socio-cognitive theories (also referred to as process and functionalist approaches), which do not explain the variability of trait expressions as error but as meaningful expressions of a stable personality system (Mischel & Shoda, 1995; Wood, Gardner, & Harms, 2015). These theories typically explain traits through a process lens: the covariation of traits is shaped by affective, cognitive, and behavioral characteristics (which potentially correspond to multiple traits) that have been reinforced by various goals, situations, and environments (Cervone, 2005; McCabe & Fleeson, 2012; Mõttus et al., in pressa).

**1.2.1. Coventional socio-cognitive theories.** One conventional socio-cognitive theory is the cognitive-affective personality system theory (CAPS; Mischel & Shoda, 1995, 2008). CAPS theory proposes that social, emotional, and cognitive systems, which are composed of unique units, mediate the interactions between behavior and situations, allowing for both stability and variability of a person's behavior across situations. Rather than defining personality from a person-centric perspective, personality can be defined through situations and how they influence behaviors. Thus, Mischel and Shoda's (1995) CAPS theory is positioned to operate without structural traits.

To make such a theory operational, they represent personality as a system of networks. Therefore, CAPS theory already implies the use of psychometric network models but also network models more generally (e.g., neural networks; Shoda, LeeTiernan, & Mischel, 2002). The cognitive-affective units of CAPS theory, for instance, are represented as an interconnected network, and are associated with a class of behaviors in a given situation (Cramer et al., 2012a). Although this network of cognitive-affective units is a relative black box (i.e., the units are not formalized), there is some evidence that personality expressions vary in their organization of certain cognitive (e.g., semantic categorizations; Christensen, Kenett, Cotter, Beaty, & Silvia, 2018) and affective (e.g., attitudes; Dalege et al., 2016) units.

In general, CAPS theory suggests that situations influence the organization of these units (and their interactions), leading to different behavioral expressions across situations (Shoda et al., 2002). Stability across situations might suggest a general behavioral tendency, while idiosyncratic behaviors might suggest effects of the situation (Perugini, Costantini, Hughes, & De Houwer, 2016). Psychometric network models offer an appealing approach to contextualize personality by examining how connectivity between behavioral expressions are affected within and across situations (e.g., how talkative a person is around friends or strangers, how anxious a person feels when facing a stressful or mundane event; Costantini & Perugini, in press; Costantini et al., 2019; Dunlop, 2015; Mõttus et al., in pressa).

One psychometric network technique called the *Fused Graphical LASSO* (FGL;

Danaher, Wang, & Witten, 2014) has the potential to directly compare how between-person behavioral nuances interact within the same situation or how within-person behavioral nuances interact in different situations (Costantini & Perugini, in press; Costantini et al., 2019). The FGL enables this comparison by allowing the estimation of two networks simultaneously. This concurrent estimation identifies edges that are shared and unique between the two networks, facilitating a direct comparison between the two networks.

On the one hand, researchers could use FGL and ecological momentary assessment (EMA) methods to identify situations that are shared between people to examine which behaviors are consistently connected, suggesting effects of the situation. Thus, the effects of situations could be explicated, elaborating on the extent to which they affect the co-occurrence of certain behaviors (Costantini & Perugini, in press). On the other hand, FGL and EMA methods could be used to identify the behavioral patterns within the same person but in different situations. Common connections between behaviors might suggest a more stable trait, while unique connections between behaviors might suggest more situation-specific interactions. The combination of these between- and within-person analyses would provide a direct examination of the conjectures proposed by Mischel and Shoda's (1995) CAPS theory. FGL can also be applied to cross-sectional data allowing group or trait comparisons to be made (e.g., females and males on social interaction characteristics; Costantini & Perugini, in press).

**1.2.2. Functionalist socio-cognitive theories.** Functionalist socio-cognitive theories differ from conventional socio-cognitive theories by operating within a trait framework. Traits, from this perspective, are formed through the previous functionality of behaviors and represent the compensatory increases (or decreases) in behaviors that have achieved (or failed to achieve) desired outcomes (Allport, 1937; DeYoung, 2015; Wood et al., 2015). Between-person differences are thus reflected by the extent to which a person seeks and values a desired outcome across situations (Perugini et al., 2016). In this regard, the structure of traits are the between-person differences in the consistency and reinforcement of

behaviors across situations (Mõttus et al., in pressa).

Wood and colleagues (2015) describe these behaviors as functional indicators, which are denoted by three classes: abilities/efficacies, expectancies, and valuations. *Abilities/efficacies* are behaviors that are a means to a desired end. A higher ability or efficacy of a behavior (e.g., telling a joke) implies a greater likelihood of producing that behavior to achieve an end (e.g., being perceived as funny). *Expectancies* are the expected effects of a person's behavior on the environment. Two people might perform a similar behavior (e.g., tell a joke) but the outcome of the behavior differs (e.g., getting a laugh or hearing crickets), leading to alternative expected effects for the same behavior. *Valuations* are the extent that a behavior leads to an experience that is desirable or valuable. Two people, for instance, might perform a similar behavior (e.g., tell a joke) and achieve the same outcome (e.g., get a laugh) but the value they place on that outcome differs and may differentially influence whether that behavior will be performed again (e.g., whether getting a laugh is an incentive to tell more jokes).

This theoretical interpretation is conceptually provided by formative latent variables models, which suggest that behaviors are the cause of interindividual traits. Thus, traits are viewed as the average composites of behavioral variables (e.g., socio-economic status being defined by occupation, education, and income; Bollen, 2002; Schmittmann et al., 2013). This also affords the interpretation that some behaviors may influence traits independently of one another, suggesting that distinct indicators explain why two traits covary. Wood and colleagues' (2015) Figure 1B explicitly demonstrates this by depicting arrows pointing from behaviors to traits rather than traits to behaviors.

From a psychometric network perspective, functional indicators are nodes and the connections between them serve as an explicit representation of which indicators overlap and form certain personality characteristics. The interpretation that these connections are mutually reinforcing lends credence to the notion that certain indicators may be directly or indirectly compensatory of one another. For example, two indicators may be positively

related, while at the same time having opposite relations with another indicator (i.e., one positive and the other negative). Furthermore, indicators that are not connected to one another are conditionally independent, suggesting they lack common underlying processes (or goals; Wood et al., 2015). Thus, psychometric networks allow an approach to directly identify certain compensatory behaviors by modeling how functional indicators relate to certain characteristics, goals, and situations (Costantini & Perugini, in press; Costantini et al., 2015b, 2019).

### 1.3. Integrative Trait Theories

Conventional and socio-cognitive theories of personality tend to emphasize between-person and within-person explanations of traits, respectively. Many modern trait theories, however, are integrative, meaning they attempt to provide both between-person and within-person explanations for the structure, process, and development of personality. In general, these theories tend to encompass conventional and socio-cognitive models, emphasizing the explanation of the whole person rather than certain aspects of personality. Thus, modern integrative theories come full circle on the historical crux of the field: understanding the whole person (Allport, 1937).

One integrative theory that builds directly from the FFM is Whole Trait Theory (WTT; Fleeson & Jayawickreme, 2015). WTT proposes that everyday trait-relevant behaviors represent states, which form density distributions. These density distributions are a person's traits, representing recurring states that have accumulated across situational experiences but also the variation in states due to contextual influences (Fleeson, 2004; Mõttus et al., in pressa). In this framework, Fleeson (2001) suggests that the density distributions are meaningful themselves and should be considered a person's whole personality rather than only one aspect (e.g., the mean) of the distributions. The density distributions, however, only represent the descriptive side of traits—what one actually does (Fleeson & Jayawickreme, 2015).

The explanatory side—what one is capable of—represents a distinct entity of traits consisting of socio-cognitive mechanisms (e.g., goals, beliefs, values, scripts, life stories) that are linked to affect and motivation, which influence the interpretation of changing situations and events. These mechanisms are viewed as the causes of the descriptive side of traits, explaining the variability and stability of the distribution of states. Thus, although distinct entities of traits, WTT suggests that the descriptive side is inextricably linked to the explanatory side by causal mechanisms. In this regard, the structure of traits is formed through causal socio-cognitive mechanisms that are described by stable distributions of states.

This theoretical perspective is conceptually provided by multilevel models. In these models, the descriptive side of traits are represented by between- and within-person variation, and the explanatory side of traits are represented by socio-cognitive mechanisms (e.g., goals) that predict moment-to-moment trait variations (i.e., states; McCabe & Fleeson, 2012). Notably, if the means of each person are factor analyzed, then five factors are typically produced (Borkenau & Ostendorf, 1998). Although multilevel models are an attractive statistical model for achieving integration, they suffer from aggregation across behaviors (even within people). This aggregation hinders any inference into the temporal interactions between behaviors, which are necessary to specify intraindividual processes.

Epskamp, Waldorp, Mõttus, and Borsboom (2018c) developed a multilevel network model (MNM) framework, which serves as an appealing alternative to standard multilevel modeling. The MNM is like standard multilevel models because it estimates fixed (between-person) and random (within-person) effects but differs because it models interactions between behaviors (e.g., self-report items, goals) rather than behavioral aggregates (e.g., traits). The MNM framework offers three types of representations: temporal (intraindividual relations from a previous time point to the next), contemporaneous (intraindividual relations within single time points), and between-subjects (interindividual relations across time points; Epskamp et al., 2018c).

These three representations offer a powerful framework for integrating intraindividual processes with interindividual structure (Baumert et al., 2017; Beck & Jackson, 2017; Breil et al., 2019; Sosnowska, Kuppens, De Fruyt, & Hofmans, 2019). The temporal representation allows causal relations to emerge, identifying which behaviors typically lead to others across time. The contemporaneous representation facilitates insight into behaviors that tend to co-occur within individuals within time points. The between-subjects representation provides inference into how the structure of these processes emerges across people. When taken together, these representations provide a rich explanation of intraindividual processes to interindividual structure and they offer a testable framework for whether traits are the emergent properties of causally coupled behaviors or are the shared influence of an underlying endogenous entity (Mõttus & Allerhand, 2017). Recent work already hints at the potential for both (Mõttus et al., in pressa; Mõttus, Epskamp, & Francis, 2017).

### 1.4. Summary

In this section, we provided a brief description of the theoretical foundations for several modern personality theories and we discussed how current psychometric models conceptually provide the basis for their explanations. After each explanation, we offered ways that psychometric networks could further inform theories that are not conceptually supported by current psychometric approaches. It's worth noting that several other modern personality theories not discussed in this section have already been influenced and informed by network models. For example, the MIMIC model (multiple indicators, multiple causes; Kievit et al., 2012) in the Cybernetic Big Five Theory (DeYoung, 2015), neural networks in the Virtual Personalities Model (Read & Miller, 2002; Read, Droutman, & Miller, 2017; Read et al., 2010), and psychometric networks in the mutualism model (Cramer et al., 2012a). Across modern personality theories, psychometric networks represent a new methodological tool to explore the validity of their claims.

## 2. Personality Measurement and the Network Approach

Despite many modern personality theories' motivation for latent variable and network models, measurement in personality research is dominated by CTT. This is likely for many reasons. One reason is that latent variable models tend to be more sophisticated, requiring additional training on behalf of the researcher to use these types of models (e.g., IRT; Borsboom, 2006). Another reason is that psychometric networks are still relatively novel and are slowly being integrated into personality research. To date, network analysis has mainly been applied in psychopathology and has only recently been applied to personality (e.g., Costantini et al., 2015a, 2015b; Cramer et al., 2012a). The most significant reason, however, is that traits, almost unequivocally, have been the unit used in the theoretical interpretations of statistical analyses (Dunlop, 2015; Mõttus, 2016).

When studies measure personality–outcome associations, they almost exclusively emphasize traits (usually the FFM). In most cases, researchers administer a personality inventory, sum up the scores for each trait, and correlate the sum scores with the outcomes. When focusing on a single trait, some studies use a latent variable (i.e., a weighted sum score) or analyze facet-level relations (but typically do not interpret their theoretical contribution; Mõttus, 2016). Therefore, personality is usually reduced to composites (or common covariances) that emphasize the whole (i.e., trait domains), with little consideration of the parts (i.e., characteristics and nuances). Emerging evidence, however, suggests that personality characteristics and nuances are as heritable and stable as domains (Mõttus, Kandler, Bleidorn, Riemann, & McCrae, 2017; Mõttus et al., in pressb) and have more predictive power than the domains they constitute (Anglim & Grant, 2016; Mõttus, Bates, Condon, Mroczek, & Revelle, 2018; Seeboth & Mõttus, 2018).

One of the main measurement advantages of the network approach is that the parts are not obscured by the whole. The network approach offers a richer representation of traits by providing a perspective of the whole through intuitive visualizations of the interactions between the parts (Bringmann & Eronen, 2018). This representation opens new avenues for

identifying influences of personality characteristics on outcomes and formalizing the role of specific elements in the personality continua and hierarchy. Mirroring current trends, psychometric networks allow researchers to emphasize the elements of traits. Using recent research, we illuminate how this perspective affords unique measurement advantages for personality research.

### 2.1. Personality–Outcome Associations

The motivation to analyze personality–outcome associations using characteristics and nuances draws from parallels with molecular genetics (Seeboth & Mõttus, 2018). In genetics, its common for complex phenotypes (e.g., neuroticism, psychiatric disorders) to be linked to hundreds and thousands of genetic variants (Visscher, Brown, McCarthy, & Yang, 2012). Similar approaches have been applied in personality research usually linking items to many different outcomes (Seeboth & Mõttus, 2018; Weiss, Gale, Batty, & Deary, 2013).

These studies suggest that relatively small effects can contribute to more accurate predictive models, highlighting their combinatorial, rather than isolated, nature. Furthermore, they implicate the complexity of phenotypes and that many personality outcomes are likely to be polynuanced (i.e., driven by a large number of personality characteristics and nuances rather than broad traits; Seeboth & Mõttus, 2018). One criticism of these studies, however, is that they tend to be atheoretical and driven solely by statistical prediction (Visscher et al., 2012). Additionally, more predictor variables may also lead to better predictions in linear and logistic regression models.

Traditional practices in personality research can shield researchers from taking such atheoretical approaches. Trait domains, for example, provide perfunctory tests for whether nuances are linked to an outcome. If a domain is not associated with an outcome, then a researcher is less likely to probe deeper into the relation. This is not to say that characteristics and nuances will not be related to the outcome; however, it prevents post-hoc digging in the data. If a domain is associated with an outcome, then it's insufficient to

suggest that the domain is the reason for the relation; instead, researchers should examine facet (and item) associations to determine whether the associations are similar across all facets (and items; Mõttus, 2016).

Statistical techniques can be used to prevent researchers from over-fitting their data when examining many personality nuances with outcomes. Cross-validation, for example, can lead to more generalizable results (Yarkoni & Westfall, 2017). Regularized regression techniques, such as the least absolute and shrinkage selection operator (LASSO; Tibshirani, 1996), are also useful for reducing over-fitting. Another appealing quality of regularization approaches is that they penalize complex models, leading to greater parsimony. When used in combination, these techniques can produce higher quality results.

Seeboth and Mõttus (2018) recently applied these statistical techniques to a large sample ($N \sim 8{,}700$) who completed 50 items from the International Personality Item Pool (Goldberg, 1999) and 40 measures of outcomes from a variety of life domains. These outcomes were then predicted using a model containing 50 individual items and a separate model containing 5 domain scores (summed totals of their respective items). Across all 40 outcomes, they found that the item models predicted as much or more variance than the domain models. And although predicted variance was relatively small (ranging from .1% to 15.5%), the item models explained, on average, 30% more variance than the domain models. Their findings suggest that personality–outcome relations are often due to specific personality characteristics rather than trait domains.

In a psychometric network, outcome variables can be included in the network. For partial correlation network approaches, such as the graphical LASSO (GLASSO; Epskamp & Fried, 2018; Friedman, Hastie, & Tibshirani, 2008), regularized partial correlations between outcomes and personality characteristics can be provided in a single illustration (e.g., Costantini & Perugini, 2016; Costantini et al., 2015b). Beyond supplying an intuitive map of relations, psychometric networks allow compensatory connections between behaviors to be revealed, facilitating a functionalist interpretation.

It's likely, for example, that many item–outcome relations in Seeboth and Mõttus's (2018) item model were shrunk to zero, suggesting that they were not contributing to the explained variance of the outcome. Several items, however, may have been connected to the outcome through other items, contributing to the variance explained in an indirect way. Indeed, the network measure *predictability*—how well a node can be predicted by the nodes it's connected to—can be used to determine the direct variance explained by variables in the network, while the connectivity between items can inform their indirect influence on the outcome (Haslbeck & Waldorp, 2018). Such interpretations are not feasibly inferred from latent variable models. In addition, latent variable models often require extensive manipulations to achieve an appropriate fit for a model this complex (if a model can even converge). A similar approach can be applied to facet–outcome relations. LNM can be used to estimate latent facets, which can then be used in a network model with outcomes, removing measurement error while providing a more nuanced picture of the associations between personalty and outcomes.

### 2.2. Personality Continua

Personality as continua is a pervasive idea in personality measurement (e.g., Likert scaling; Tay & Jebb, 2018). This notion stems from interindividual differences where one person, relative to another, endorses more behaviors associated with a certain trait. In measuring a trait's continuum, item scores are typically summed across a scale, which imposes interindividual variability. Although sum scores provide a continuous measure, they by no means imply that a trait's continuum is being measured. Because most personality scales were developed using CTT, they were likely constructed to have high internal consistency by retaining items that had moderate to high inter-item correlations (Chernyshenko et al., 2007; Simms, 2008). This approach, however, has led current scales to measure the midpoint well but the extremes poorly, limiting their coverage of the personality continua (Carter, Miller, & Widiger, 2018).

Measurement of a trait's continuum starts with defining both ends of the continuum. Although this seems obvious, it's often overlooked in scale construction (Tay & Jebb, 2018). The labels of the FFM, for example, all represent the high normal end of the continua and measures are typically constructed with this pole in mind. This is reflected in historical and modern accounts of a "healthy" personality where attributes are all at the positive end of the continua (Bleidorn et al., in press). Indeed, most research investigating abnormal personalities (e.g., personality disorders) have mainly emphasized the negative poles of normal personality (Krueger, Derringer, Markon, Watson, & Skodol, 2012; Widiger et al., 2019). This has led to an assumption that the relationship between personality and psychopathology is linear, despite evidence for curvilinear relationships (Carter et al., 2018). Some researchers, however, have demonstrated that extremes at the positive pole of personality are also related to psychopathological outcomes (Blain, Longenecker, Grazioplene, Klimes-Dougan, & DeYoung, 2019; Widiger & Crego, in press; Widiger, Gore, Crego, Rojas, & Oltmanns, 2017).

Renewed interest in the measurement of the personality continua has largely been driven by the need for both ends of the extremes to be integrated into normal and abnormal personality constructs (Carter et al., 2018; Durbin & Hicks, 2014; Widiger et al., 2019). There is ample evidence demonstrating that normal personality provides a fruitful framework for psychopathology, especially for personality disorders (Saulsman & Page, 2004; Widiger & Simonsen, 2005; Widiger et al., 2017). Much of this evidence has been derived from correlational and factor analytic comparisons of normal and abnormal personality inventories (Saulsman & Page, 2004; Watson, Clark, & Chmielewski, 2008; Watson, Ellickson-Larew, Stanton, & Levin-Aspenson, 2016; Wright & Simms, 2014). Markon and colleagues (2005), for example, factor analyzed a meta-analytically derived correlation matrix as well as new data sets and found that normal and abnormal personality scales could be integrated into a common hierarchical structure. This study, among others (e.g., Clark & Livesley, 2002; Watson et al., 2008; Widiger & Simonsen, 2005), has inspired several five-factor personality

disorder inventories, including the Personality Inventory for DSM-5 (Krueger et al., 2012; Widiger et al., 2017). Although these studies are useful for integrating normal and abnormal personality hierarchies, they have not decisively demonstrated that abnormal personality lies at the poles of the personality continua.

Some researchers have turned to IRT to evaluate whether abnormal personalities are more extreme variants of normal personality traits (Samuel, Simms, Clark, Livesley, & Widiger, 2010; Stepp et al., 2012; Walton, Roberts, Krueger, Blonigen, & Hicks, 2008). These studies have typically found that normal and abnormal personality inventories strongly overlap but that abnormal inventories tend to measure the negative end of the personality continua. Ideal-point scoring models have been proposed as an approach to develop measures that capture both ends of the personality continua (Carter et al., 2018; Chernyshenko et al., 2007). Ideal-point IRT methods allow for people to disagree with an item because they are either higher or lower in the trait level expressed by the item rather than just lower. For instance, a person might disagree with a moderate neuroticism item such as "I sometimes can't help worrying about the little things" because they *always* worry about the little things. There are a growing number of studies that demonstrate that ideal-point models fit personality data better than dominance models but also provide more information about items' coverage of a scale's continuum, leading to greater construct and criterion-related validity (Chernyshenko et al., 2007; Stark, Chernyshenko, Drasgow, & Williams, 2006).

Beyond domain-level poles, few studies have examined the continua of characteristics. Chernyshenko and colleagues (2007), for example, constructed an orderliness scale (a characteristic of conscientiousness), which was used as their vehicle to demonstrate the effectiveness of ideal-point process models. When compared to a scale developed using traditional CTT rules of thumb, they found the ideal-point process model had better criterion-related validity and more information across the trait's continuum.

Similarly, few studies have connected the continua of normal and abnormal personality at the facet-level (Samuel & Widiger, 2008; Thomas et al., 2013; Watson, Stasik,

Ellickson-Larew, & Stanton, 2015). Samuel and colleagues (2008) performed a 16 study meta-analysis that identified the effect sizes between FFM facets and personality disorders and revealed that each personality disorder has the unique profile of FFM facet associations. Some disorders fit within one domain of facets while others spread across several domains and their facets. Their study underscores the importance of going beyond trait domains to establish normal and abnormal personality continua at lower levels of the hierarchy.

Psychometric networks provide an approach to integrate normal and abnormal personality beyond trait domains. Unlike latent variable and ideal-point models, psychometric networks allow specific inferences into where the connections between normal and abnormal characteristics (and nuances) blur. Comorbidity in the psychopathological literature provides a convenient template for making these inferences. Cramer, Waldrop, van der Mass, and Borsboom (2010), for example, mapped the symptom space of two commonly comorbid disorders: major depressive disorder (MDD) and generalized anxiety disorder (GAD). They illustrate that some symptoms of MDD and GAD are shared and form "bridges" between the two disorders, smearing the lines where one disorder begins and the other ends.

Just as symptoms of MDD blur into symptoms of GAD, normal personality characteristics can be expected to blur into abnormal characteristics. In a normal and abnormal network of personality characteristics, researchers can identify normal or "healthy" by characteristics that have few or no connections to abnormal characteristics. Similarly, abnormal characteristics that are not connected to normal characteristics are likely representative of the most extreme end of the continuum. The intermediate connections are the blurred boundaries between normal and abnormal personality. This representation, combined with ideal-point models, would allow researchers to better define the constructs they're measuring and identify areas where the personality continua is insufficient.

### 2.3. Personality Taxonomy

Scientific taxonomies seek to provide a descriptive model of organized concepts that simplifies a vast number of variables into overarching domains. The general consensus among personality researchers is that personality traits are hierarchically organized at different levels of breadth and depth (John & Srivastava, 1999; McCrae & Costa, 2008). Usually, trait domains are decomposed into facets, which are further broken down into items (McCrae & Costa, 1987, 2008). More recently, metatraits (above traits) and aspects (between traits and facets) were added to the hierarchy (DeYoung, 2015; DeYoung, Peterson, & Higgins, 2002; DeYoung, Quilty, & Peterson, 2007; Digman, 1997). Many personality researchers agree that at least five or six trait domains represent the majority of interindividual variation (but see Block, 1995; McAdams, 1992). Below these domains, however, there is little consensus about the content and number of facets (Connelly, Ones, & Chernyshenko, 2014a; Simms, 2009).

One reason for this variety is that scale developers have typically focused on developing instruments at the domain-level, with less attention devoted to lower-order structures. In 1999, Goldberg made a critical call to "[change] the way that we construct new measures of personality characteristics" (p. 7); yet, only a few inventories since then have heeded his plea. One notable example is the Big Five Aspects Scale (BFAS; DeYoung et al., 2007). DeYoung and colleagues (2007) factor analyzed the facets (75 in total) of two existing inventories—the NEO PI-R and AB5C-IPIP (Goldberg, 1999)—and derived two distinct yet related aspects for each Big Five domain. These distinctions within each domain have shown considerable convergent and divergent validity, with many aspects producing differential predictions with outcomes (e.g., Openness and Intellect with creative achievement in the arts and sciences, respectively; Kaufman et al., 2016). One constraint of the BFAS is that it does not differentiate facets within its aspects.

More recently, two new assessment instruments were developed to meet the growing focus on facets (Soto & John, 2017; Watson, Nus, & Wu, 2017). Soto and John (2017) recently developed a revision of their popular Big Five Inventory (BFI; John, Donahue, &

Kentle, 1991). The initial version of the BFI was designed to measure prototypical features of each Big Five domain and only later were facets derived (Soto & John, 2009). Soto and John's revised version, BFI-2, explicitly emphasized the development of lower-order facets and aimed to establish a balanced bandwidth (breadth) and fidelity (depth) at both domain and facet levels. A key notion in the development of their facets was first selecting factor-pure facets (i.e., central to the domain and independent of other domains) to provide a foundation for the domain. After, they selected two complementary facets (i.e., related to the domain but also other domains) to broaden each domain, providing greater conceptual continuity with other domains. Their final inventory resulted in a 60-item questionnaire that measured 15 facets (three per trait), with 4 items per facet (twelve per trait). The revised version's facets demonstrated substantial increases in predictive power over the initial version's ad hoc facets.

Watson and colleagues (2017) introduced another hierarchical inventory called the Faceted Inventory of the Five-Factor Model (FI-FFM; Simms, 2009). Construction of the FI-FFM proceeded over five stages of data collection and analysis, beginning with the development of specific scales and ending with refinement of existing scales and the creation of new scales. A notable feature of the FI-FFM is that the facets of the five domains are measured asymmetrically (i.e., each domain has a different number of facets and items). The advantage of such an approach is that the inventory is assembled based on each domain's conceptual complexity; thus, it's not limited to artificial constraints of an equal number of facets per domain. They developed 22 facets in total, which were selected based on an over-inclusive review of existing literature and integrative trait structure work (Roberts, Chernyshenko, Stark, & Goldberg, 2005; Simms, 2009). Another notable feature is that they developed abnormal personality facets alongside their other facets; however, these did not fit cleanly into the structural analyses and were subsequently dropped. Despite the lack of inclusion, this represents one of the first attempts to develop a personality scale that included both normal and abnormal characteristics.

Other studies have taken a more specific approach to facet development, focusing on singular domains and synthesizing their facets to derive a clearer consensus of the domain's facets (Connelly, Ones, Davies, & Birkland, 2014b; Roberts et al., 2005; Woo et al., 2014). Two of these studies concentrated on the openness to experience domain, which is perhaps the most contentious trait of the Big Five (Connelly et al., 2014a). Connelly and colleagues (2014b) performed a theoretical sort across 85 scales related to openness to experience and meta-analyzed their relations to other trait domains. They identified 11 facets and distinguished whether facets were pure (i.e., not related to any other domain; 4 facets) or compound (i.e., related to at least one other domain; 7 facets). In a different study, Woo and colleagues (2014) factor analyzed 36 scales related to openness to experience and derived a hierarchical measure that included 6 facets and 2 aspects, which were conceptually in line with DeYoung et al.'s (2007) Openness and Intellect aspects. Roberts and colleagues (2005) completed similar analyses on the trait conscientiousness, finding 6 facets (3 pure and 3 compound). A common thread among these studies is that the facet structure of a given domain was analyzed at the facet-level. One limitation of this is that facets are taken at their conceptual face value.

A simple structure perspective on facets suggests that the items of a facet should align with their own facet; however, this is not always the case (Christensen, Cotter, & Silvia, 2018a). Christensen and colleagues (2018a), for example, used a psychometric network approach to construct a network of four different inventories of openness to experience to further clarify the work of Connelly et al. (2014b) and Woo et al. (2014). Instead of a network of facets, nodes in the network were items, which allowed each item to freely associate with items outside of their intended facets. Using a community detection algorithm, which detects clusters (dimensions) in the network, they found 3 aspects and 10 facets. Two of the three aspects aligned with DeYoung et al. (2007) and Woo et al.'s (2014) aspects (Openness and Intellect), with an additional novel aspect being uncovered (Open-Mindedness). All ten facets conceptually aligned with the eleven facets derived from

Connelly and colleagues' (2014b) theoretical sort. They also found that most items aligned with their intended facets; however, there were several instances where items associated outside of their intended facets. Items in the HEXACO-100's unconventionality facet, for instance, sorted into 3 different facets, involving all 3 aspects. Christensen and colleagues' (2018a) study is an example of how psychometric networks can be used to examine the hierarchy embedded within the connections between items.

The emergent hierarchy of psychometric networks offers a couple advantages over other psychometric approaches for examining personality's taxonomy. One advantage is that the hierarchical resolution or distinction between one level of the hierarchy (e.g., facets) and the next (e.g., aspects or domains) is less discrete than a latent variable model. Items, for example, can be shown associating with one another but also with items of other facets (see Figure 1; Christensen et al., 2018a). In contrast, a traditional latent variable model only allows associations between the highest level being measured. And although the residuals of the items can be correlated, this is usually not done and adds many more parameters, leading to potential problems with identifying a model that converges (but see Herrmann & Pfister, 2013). Moreover, the levels in-between items and traits are not shown to associate amongst themselves; thus, information is being lost in every layer of the hierarchy[1] (John & Srivastava, 1999). In this way, traditional latent variable models have a discrete hierarchical resolution (i.e., simple structure; but see ESEM; Asparouhov and Muthén (2009)), while psychometric network models have a continuous hierarchical resolution (i.e., complex structure).

[1] It's important to note that latent variable models are confirmatory and psychometric network models are exploratory. Exploratory factor analysis (EFA) models do provide information about the relations between variables through their factor loadings (and correlations between factors). A notable difference between EFA models and psychometric network models, however, is that the former is a full model while the latter is a reduced model (i.e., some variables are conditionally independent of one another).

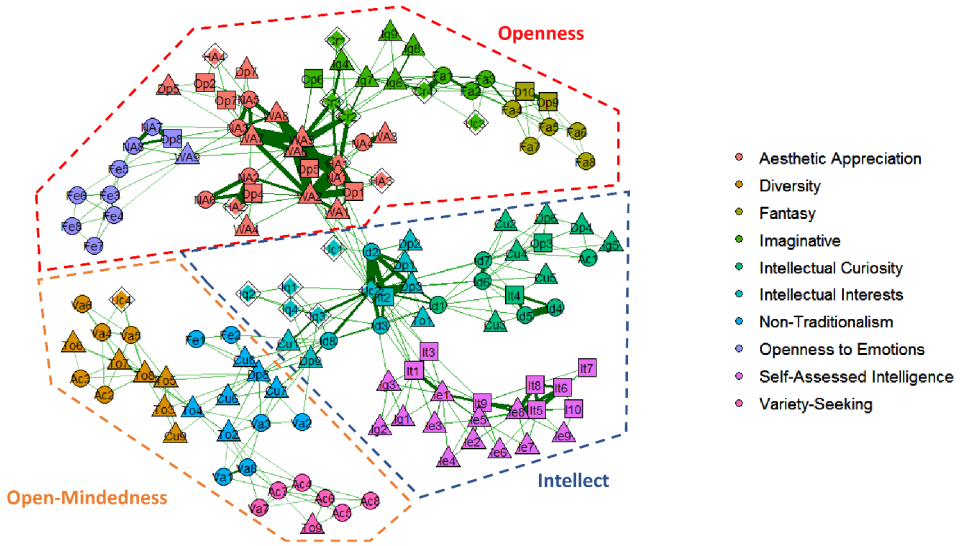


*Figure 1*. Openness to experience network from Christensen, Cotter, and Silvia (2018a). The nodes represent items and their colors are their respective facets. The dashed lines represent the aspects of openness (red), intellect (blue), and open-mindedness (orange). The entire network represents the trait domain of openness to experience.

Another advantage that psychometric network models provide is the concept of centrality (Costantini et al., 2015a; Cramer et al., 2012a). Because items are allowed to associate with one another, the notion of which facets are more central (pure) or peripheral (compound) becomes increasingly salient. In the studies discussed above, factor analytic models were used to assess whether a facet indicated a single domain or if it loaded onto other domains. It's unclear, however, what items in particular were leading to these associations and how exactly these items were associating within their own facet compared to others. Psychometric networks provide an approach to disentangle these associations. In a psychometric network, pure facets would appear closer to the center of the network (in a single domain network) and would be mainly connected to other facets associated with the respective domain (in a multiple domain network). In contrast, compound facets would appear to be more peripheral in the network (in a single domain network) and would have comparable connections within and outside of the respective domain (in a multiple domain

network). Thus, networks are able to identify exactly which items (and facets) are leading to these inter-facet (and inter-domain) associations, which can inform scale developers on whether to include certain items based on the construct they intend to measure.

### 2.4. Summary

The main advantage of psychometric network models is that they allow a representation of relations between the lower-order components of traits (e.g., items and facets). This perspective affords intuitive interpretations of personality–outcome relations, while also permitting a continuous hierarchical resolution—one *can* see the forest for the trees. In addition, constructs in networks appear to be fuzzy rather than discrete, facilitating a mereological interpretation that is better aligned with the boundaries between normal and abnormal personality (Cramer et al., 2012b). Importantly, researchers stand to gain the most when combining multiple psychometric methods. Therefore, the measurement of personality can only benefit from the integration of psychometric networks with other psychometric approaches.

To this point, we've shown that psychometric networks can inform modern personality theories and measurement in ways that other psychometric approaches can not. A fundamental constraint on the potential for psychometric networks to advance personality theory and measurement, however, is that current assessment instruments are based on CTT notions. If psychometric networks are to advance personality theory and measurement in a meaningful way, then new assessment instruments must be developed from the perspective that it provides. Indeed, if researchers are to take the perspective that personality traits emerge from causally coupled components, then it's essential to develop new measures that equivalently express them that way (Hallquist et al., 2019).

## 3. A Roadmap for Scale Development using Psychometric Networks

To date, psychometric network models have mainly been used as a novel measurement tool, which has led to an alternative account on the formation of traits. This account offers a

fruitful framework to integrate modern personality theories and affords several measurement advantages over other psychometric approaches. Therefore, psychometric networks are a promising approach for the measurement of personality (e.g., Costantini & Perugini, in press; Mõttus & Allerhand, 2017). When it comes to psychometric assessment, however, the scope of psychometric networks has been far more limited (e.g., dimension reduction methods; Golino & Epskamp, 2017).

So far, the results in the area of dimension identification appear promising and suggest that psychometric network analysis has some potential as a method for constructing personality assessments (Golino & Demetriou, 2017; Golino & Epskamp, 2017; Golino et al., 2018). Some researchers have demonstrated that certain network models (e.g., the Ising model; van Borkulo et al., 2014) are mathematically equivalent to models already used for the construction and validation of psychometric assessments (e.g., multidimensional IRT, MIRT; Epskamp et al., 2018b; Marsman et al., 2018). Others have associated the theoretical interpretations of latent variable models to network models, suggesting they may be more alike than different (Bringmann et al., 2018; Kruis & Maris, 2016). Despite these connections, research has yet to examine whether network analysis can be used as a tool for the construction of a psychometrically reliable and valid assessment instrument (Christensen, Cotter, Silvia, & Benedek, 2018b).

The aim of this section is to take the initial steps toward a formalized framework for the use of psychometric network models in scale development. To achieve this aim, this section is divided into four parts. First, we introduce psychometric network measures and techniques. Second, we connect the network measures to measures used in traditional psychometric approaches using a small-scale simulation. Third, we contrast traditional scale development guidelines with new guidelines developed from the perspectives provided by the network approach. The main focus of these guidelines will be on the evaluation of the item pool, which is arguably the most important stage in scale development. Finally, we discuss areas where future research is necessary to further establish this framework.

### 3.1. Network Measures and Techniques

A critical factor in psychometric network analysis is estimation. Network estimation is necessary because the relations between psychological variables (e.g., self-report items) are typically not known. The most popular network estimation approach in psychometric network literature is the GLASSO (Friedman et al., 2008), which shrinks edges (via regularization) and sets small edges to zero, producing a sparse and parsimonious model. The GLASSO estimates a Gaussian Graphical Model (GGM; Lauritzen, 1996), where edges represent partial correlations between variables after conditioning on all other variables in the network. Notably, there are other network estimation methods, each of which have different downstream effects on the structure and measures of a network (e.g., Christensen, Kenett, Aste, Silvia, & Kwapil, 2018c; Williams, 2018; Williams & Rast, 2018). In general, the network measures and techniques presented in this section yield similar interpretations across network estimation methods.

**3.1.1. Node-wise measures.** Centrality is a key concept in network analysis and is the most commonly applied network measure in the literature (Bringmann et al., 2018; Cramer et al., 2012a). Centrality is typically applied to individual nodes (Bringmann & Eronen, 2018; Costantini et al., 2015a) but can also be extended to communities in the network (Blanken et al., 2018; Christensen, 2018; Giscard & Wilson, 2018). There are three measures of node centrality that are most commonly applied in the psychometric network literature: betweenness, closeness, and node strength (see Table 1 for descriptions).

Table 1
*Network Measure Terminology*

| Measure | Description |
|---|---|
| **Node-wise Centrality** | |
| Betweenness (BC) | the relative number of times a node is used on the shortest path from one node to another |
| Closeness (LC) | the (inverse of the) distance a node is from the center of the network (larger values = closer to center) |
| Node strength (NS) | the sum of the weights (e.g., correlations) connected to each node |
| Hybrid (HC) | the overall centrality of each node based on the rank-order composite of several weighted (correlations = edges) and unweighted (all edges = 1) node-wise centrality measures (BC, LC, NS, and eigenvector centrality; EC) |
| **Community Centrality** | |
| Community Closeness (CCC) | the (inverse of the) distance a community is from the center of the network (larger values = closer to center) |
| Community Eigenvector (CEC) | the indirect and direct connections of a community to the rest of the network |

Node centrality measures were initially interpreted as the relative influence of a node in the network, an interpretation that was likely derived from social networks (Bringmann et al., 2018). A highly central node in a psychopathological symptom network, for example, was thought to have a strong influence over other nodes in the network, representing a potential target for clinical interventions (Borsboom & Cramer, 2013). More recent research, however, has questioned this interpretation, citing differences in social and psychopathological network estimation—that is, connections between people are known, whereas connections between symptoms are not (Bringmann et al., 2018). Network estimation also brought considerable inconsistency to some centrality measures. Betweenness and closeness centrality, for example, have consistently been shown to be unstable, leading many researchers to abandon these measures (Christensen et al., 2018c; Epskamp, Borsboom, & Fried, 2018a; Forbes, Wright, Markon, & Krueger, 2017). Indeed, in a recent simulation study, betweenness and closeness

centrality were found to be unstable due to sampling variability (Hallquist et al., 2019).

Node strength, however, has demonstrated considerable consistency and is still frequently used in the psychometric network literature (Epskamp et al., 2018a). Some researchers have hypothesized that node strength might be related to factor loadings (Mõttus & Allerhand, 2017). Consistent with this hypothesis, Hallquist and colleagues' (2019) found that node strength was roughly redundant with factor loadings. This finding not only helps establish a link between network measures and traditional psychometric measures but also helps clarify how node strength should be interpreted (Bringmann et al., 2018). In Hallquist et al.'s (2019) simulation, node strength represented a combination of dominant and cross-factor loadings, suggesting that it should be used in the context of each dimension.

One novel measure that attempts to provide relatively similar information as betweenness and closeness centrality is the hybrid centrality (Table 1; Christensen et al., 2018c; Pozzi, Di Matteo, & Aste, 2013). The hybrid centrality quantifies the overall "centralness" of nodes in the network. This is because it quantifies each node's relative distance from other nodes (BC and LC) as well as its direct (NS and EC) and indirect (EC) connections to other nodes. Recent research has demonstrated that it has predictive validity (Christensen et al., 2018c), is relatively stable (Christensen et al., 2018b), and is useful for determining labels of factors (Christensen et al., 2018a). Although this measure has not achieved widespread adoption in the psychometric network literature, it's a promising alternative to the unstable betweenness and closeness centralities.

**3.1.2. Dimension identification.** In psychometric networks, the main methods used to assess dimensionality are called community detection algorithms. These algorithms identify the number of communities (i.e., dimensions) in the network by maximizing the connections within a set of nodes, while minimizing the connections from the same set of nodes to other sets of nodes in the network. The most extensive work on dimensionality in the psychometric network literature has been with a technique called *exploratory graph analysis* (Golino & Epskamp, 2017).

Exploratory graph analysis first constructs a network using one of two network estimation methods: the GLASSO or the triangulated maximally filtered graph (TMFG; Golino et al., 2018; Massara, Di Matteo, & Aste, 2016, for descriptions of both methods see SI, Dimension Identification Methods). Then, the technique applies a weighted community detection algorithm (i.e., walktrap; Pons & Latapy, 2006, for more details see SI, Community Detection) to determine the number and content of communities in the network. So far, simulation studies have shown that exploratory graph analysis with GLASSO estimation has comparable or better accuracy at identifying the number of dimensions than traditional techniques (Golino & Demetriou, 2017; Golino & Epskamp, 2017; Golino et al., 2018). Hereafter, exploratory graph analysis with GLASSO will be referred to as EGA and exploratory graph analysis with TMFG will be referred to as EGAtmfg.

In a recent simulation study, Golino and colleagues (2018) found that EGA and EGAtmfg were comparable to the best parallel analysis (PA) with resampling techniques (principal components analysis; PApca, and principal axis factoring; PApaf). Importantly, the two exploratory graph analysis methods complemented each other's weaknesses. EGA was the most accurate of all methods examined based on percent correct (PC; correct identification of the number of dimensions) and performed best with moderate to high factor loadings and sample sizes. Notably, however, PApaf, PApca, and EGAtmfg all had better performance based on mean bias error (MBE; bias for under- and over-factoring) and mean absolute error (MAE; general bias from population value). When both exploratory graph analysis methods agreed on the number of dimensions (63% of the time), their accuracy was very high (93%, MAE = 0.11). When they disagreed, their accuracy was poor (EGA = 35.4%, MAE = 8.08 and EGAtmfg = 16.8%, MAE = 1.79). In these cases they disgreed, they had opposite biases: EGA tended to over-factor and EGAtmfg tended to under-factor.

It's clear that the exploratory graph analysis techniques perform well for identifying dimensions and often do as well as traditional dimension reduction methods. But they also possess a couple advantages over traditional methods. First, the number and content of

dimensions is deterministic: Researchers do not have to make arbitrary decisions about cut-offs on scree plots to decide the number of dimensions nor sort through factor loadings to decide on item placement (Flora & Flake, 2017; Reise, Waller, & Comrey, 2000; Zwick & Velicer, 1986). Second, the connections between variables are depicted, allowing for a continuous hierarchical resolution. Connections between items, for example, are displayed, with clusters of facets and aspects emerging from them (e.g., Figure 1). This visualization provides the researcher with useful information about how the dimensions are associated with one another as well as how central some dimensions are relative to others.

**3.1.3. Community centrality measures.** For psychometric networks, a dimension's centrality is immediately evident and should have a substantive role in scale development. For instance, facets that were visually more central in Christensen, Cotter, and Silvia's (2018a) network of openness to experience had higher correlations with the overall domain than less central facets. Relying on visual inspection, however, is qualitative and potentially prone to misrepresentation (e.g., positioning of nodes is stochastic and may not be related to actual centrality; Forbes et al., 2017).

So far, only a few measures have been developed to examine the relative position of communities in the network (Christensen, 2018; Everett & Borgatti, 1999; Giscard & Wilson, 2018). Most commonly, researchers have taken the mean or sum of node-wise centrality measures (Bell & O'Driscoll, 2018; Everett & Borgatti, 1999), which may not perform as well as community-based measures (Giscard & Wilson, 2018). Community centrality measures offer a quantitative approach to determine how central certain facets are relative to others.

Two recently developed measures are suitable candidates for measuring the centrality of communities: community closeness centrality (CCC; Christensen, 2018) and community eigenvector centrality (CEC; Giscard & Wilson, 2018). The CCC is an extension of the node-wise closeness centrality measure and is computed in a similar way (Table 1). Similarly, the CEC is an extension of the node-wise eigenvector centrality—the largest eigenvalue of an adjacency matrix, quantifying the weighted sum of a node's direct and indirect connections

(Bonacich, 1972). Giscard and Wilson (2018) generalized the node-wise eigenvector centrality to any sub-network—a subset network of the overall network. Thus, when each community is taken as a sub-network, their CEC can be computed.

Though both measures quantify the centrality of a community, they have slightly different interpretations. The CCC quantifies the relative central position of the community in the network, while the CEC quantifies the central position of the community based on its connections to other nodes and communities in the network. Therefore, they quantify separate but related aspects of a community's centrality in the network.

### 3.2. Connecting Network Measures to Traditional Psychometrics

An important step for using psychometric networks as a scale development tool is connecting network measures to established psychometric measures. To establish these connections, we performed a small-scale simulation to assess the performance of exploratory graph analysis with ordinal data and whether certain node and community centralities are associated with traditional psychometric measures. This simulation seeks to build on Golino et al. (2018) and Hallquist et al.'s (2019) simulation studies. Golino and colleagues' study compared the accuracy of dimension identification of both exploratory graph analysis methods and other more traditional dimension reduction methods in dichotomous data. Hallquist and colleagues' first study identified the correspondence between node centrality measures and factor loadings using continuous variables.

**3.2.1. Small-scale simulation.** Importantly, the simulation presented in the Supplementary Information differs from Golino et al. (2018) and Hallquist et al.'s (2019) studies in several key ways. First, the simulated datasets in this study were specifically designed to correspond with conditions that are typically found in traditional personality scales (e.g., many dimensions, 5-point Likert scaling, skewed responding) and scale development settings (e.g., wide ranging factor loadings and correlations between factors). In total, there were 2700 simulated datasets (100 datasets per manipulated condition) across 3

manipulated conditions—sample size (500, 1000, and 2500), number of dimensions (2, 4, and 6), and number of variables per dimension (6, 8, and 10). Factor loadings and correlations between factors were allowed to vary randomly between .30 and .70. For full descriptions of our design, methods, measures, and results see the Supplementary Information.

Second, we aimed to compare the dimension identification accuracy of the exploratory graph analysis methods with state-of-the-art dimension reduction approaches: PApca and PApaf. This simulation differs from a previous comparison of these methods by categorizing continuous variables into 5 categories (polytomous) rather than 2 (dichotomous; Golino et al., 2018). Furthermore, factor loadings and correlations between factors were allowed to randomly vary in this study, while they were manipulated conditions in Golino and colleagues' (2018) study. Notably, factor loadings and correlations between factors had the largest effects across all methods' on performance (PC and MAE) in their simulation. To our knowledge, this is the first simulation to examine the dimension identification performance of exploratory graph analysis with ordinal variables.

Third, we computed several network and traditional psychometric measures for comparison. Because betweenness and closeness centrality reflected relatively poor performance in Hallquist et al.'s (2019) study, they were not included in this study. Instead, the hybrid centrality measure was used. Both node strength and hybrid centrality were correlated with item-scale correlations and CFA factor loadings. We note that factor loadings and item-scale correlations are strongly related to one another. Specifically, CFA factors represent idealized sum-scores because the error-variance that is not associated with the latent factor is removed. Therefore, item-scale correlations and factor loadings are largely redundant but reflect different variance contributions (i.e., without and with error, respectively).

We also computed CCC and CEC, which were compared to factor loadings onto a general factor (hereafter referred to as general loadings) and the sum-score of factors correlated with the sum-score of the general factor (hereafter referred to as scale-measure

correlations). Again, general loadings represent idealized sum-scores of scale-measure correlations. The correlations between the traditional psychometric measures were used as benchmarks for the comparison of the network measures to the traditional measures.

**3.2.2. Summary of simulation results.** The dimension identification results paralleled Golino et al.'s (2018) simulation study with the finding that EGA was the most accurate dimension identification method (DIM) overall (PC = 77.0% and MAE = 0.31). Unlike their results, EGAtmfg (PC = 76.0% and MAE = 0.31) was the second most accurate followed by PApca (PC = 69.8% and MAE = 0.34) and PApaf (PC = 59.3% and MAE = 0.91). Notably, despite PApaf's low overall accuracy, it was the most accurate (PC = 88.4% and MAE = 0.12) DIM when sample size was large (N = 2500). Across all DIMs, the number of dimensions had the largest effect on accuracy (all moderate effect sizes). Notably, these effects should be interpreted with caution as factor loadings and correlations between factors have shown larger effects in past work (Golino et al., 2018).

There was a general tendency to under-factor for all methods except for PApaf, which tended to over-factor. For normalized mutual information (NMI) or the accuracy of item placement, EGA performed the best (NMI = 0.94) followed by PApaf and PApca (both NMI's = 0.93). Despite EGAtmfg having the second best overall accuracy for identifying the number of dimensions, it had the lowest NMI (.90). This suggests that although EGAtmfg is accurate, it's items are not always being placed in the correct dimensions (relative to other DIMs). Regardless, both exploratory graph analysis techniques performed as well as or better than the more traditional DIMs across all accuracy and bias measures. Mirroring Golino and colleagues' (2018) finding, when EGA and EGAtmfg agree on the number of dimensions, they are extremely accurate (PC = 90.4% and MAE = 0.12). When they disagree, both EGA (PC = 39.2% and MAE = 0.83) and EGAtmfg (PC = 35.1% and MAE = .86) have very low accuracy.

The centrality measures (both community and node-wise) were all strongly related to their corresponding traditional psychometric measures. For the community centrality

measures, CCC had similar effect sizes for general loadings ($\bar{r}_{EGA}$ = .69 and $\bar{r}_{EGAtmfg}$ = .58) and scale-measure correlations ($\bar{r}_{EGA}$ = .68 and $\bar{r}_{EGAtmfg}$ = .57), while CEC was more closely related to scale-measure correlations ($\bar{r}_{EGA}$ = .74 and $\bar{r}_{EGAtmfg}$ = .70) than general loadings ($\bar{r}_{EGA}$ = .56 and $\bar{r}_{EGAtmfg}$ = .53). Across these measures, EGA had much higher average correlations than EGAtmfg. Notably, these correlations were on par with the relationship between general loadings and scale-measure correlations ($\bar{r}$ = .70).

For the node-wise measures, both the hybrid centrality and node strength were strongly associated with factor loadings and item-scale correlations. In general, node strength had higher correlations with factor loadings ($\bar{r}$ = .87) and item-scale correlations ($\bar{r}$ = .86) than the hybrid centrality ($\bar{r}$ = .82 and $\bar{r}$ = .78, respectively). Notably, for both measures, EGAtmfg had slightly higher average correlations than EGA. Despite the differences between Hallquist et al. (2019) and our simulation, our results largely corroborate their findings, suggesting that node strength centrality is relatively redundant with factor loadings. Our simulation also extends their findings to suggest that node strength is redundant with item-scale correlations.

In addition, our results demonstrate that community centrality measures can provide similar information about a dimension's latent (general loadings) and direct (scale-measure correlations) contribution to the overall construct being measured. Finally, this was the first simulation, to our knowledge, to investigate the performance of exploratory graph analysis with ordinal variables. Although the number of categories tends to be a minor aspect of dimension identification performance (Garrido, Abad, & Ponsoda, 2011, 2013), the evidence provided by this simulation suggests that both exploratory graph analysis techniques performed well.

### 3.3. Psychometric Network Scale Development Guidelines

The associations between traditional psychometric measures and centrality measures form a bridge between familiar measures of CTT and factor analysis and the less familiar

measures of psychometric networks. This bridge serves as a foundation for establishing how psychometric networks can be used as a scale development tool. In general, psychometric networks offer two substantive contributions for how scales should be developed. First, the interactions between items are the focus rather than the scales themselves. Second, the intuitive representations of these interactions provide a multidimensional perspective of how items are positioned with respect to the rest of the scale. This emphasis on the items, their interactions, and their function in a multidimensional context fundamentally changes how items should be selected. More or less, traditional guidelines still hold as a tried-and-true framework for scale development; however, psychometric networks offer an opportunity to expand this framework.

**3.3.1. Dimensionality assessment.** Dimensionality assessment is an integral step of verifying the structural validity of a scale. Often this step is performed later in the item evaluation stage; however, item selection measures (e.g., item-scale correlations) hinge on what dimension items belong to. Personality scales are usually multidimensional and multifaceted—for every inventory, there typically are five or six trait domains and for every domain, there are several facets (ranging from two to nine). In traditional psychometrics, factor analytic techniques, such as PCA and EFA, are the most common method to assess the dimensionality of an inventory or domain. These methods require the researcher to carefully inspect and make decisions about item placement. In contrast, exploratory graph analysis is deterministic, meaning it decides on item placement without the researcher's direction.

Exploratory graph analysis is the vehicle for psychometric network dimensionality assessment. Although the allocation of items into dimensions is deterministic, researchers should still verify their theoretical consistency—theory should always guide the decision for the number and content of dimensions. A benefit of this determinism is that exploratory graph analysis can be used as tool to evaluate whether researchers are measuring the dimensions they intended to measure, without a priori direction from the researcher. Moreover, because the technique is exploratory, replicating the content and number of

dimensions provides "a more rigorous replication test than confirmatory analysis" (Saucier & Goldberg, 1996, p. 35; e.g., Waller, DeYoung, & Bouchard, 2016). Thus, exploratory graph analysis can serve as a test for whether items are sorting into their expected facets or whether facets that were specified as central are actually more central relative to other facets.

Furthermore, exploratory graph analysis can be combined with LNM to determine higher-order structures of a scale (e.g., aspects). First, exploratory graph analysis can be applied, identifying the facet structure of a construct. After, LNM can be used to model the latent variables of these facets. Then, a community detection algorithm (e.g., walktrap) can be applied to the LNM, revealing a higher-order structure of the facets. This approach allows a researcher to deterministically identify the hierarchical structure of their measure, testing whether this structure is being measured as intended.

When using exploratory graph analysis, researchers are recommended to apply both methods to examine the number of dimensions identified (Golino et al., 2018). If both methods return the same number of dimensions, then researchers can be confident that they've estimated the correct number of dimensions. If they return a different number of dimensions, then researchers should consider the biases of both methods (e.g., Golino et al., 2018). Measures of fit and cross-validaton techniques (e.g., Kan et al., 2019) could offer some help to determine the number of dimensions that fit the data best. In addition, a recently developed approach called *bootstrap exploratory graph analysis* (bootEGA; Christensen & Golino, 2019) can be used to estimate the stability and generalizability of these results, including the dimensional structure of a typical network from a sampling distribution of networks.

**3.3.2. Dimension centrality.** In traditional scale development, most developers evenly select items for each facet and each domain (e.g., NEO PI-R; 8 items per facet and 6 facets per domain). This view likely stems from CTT and the use of sum scores, which suggests that even item selection ensures that each facet is similarly scaled and has equal contributions to the overall score. Similar ideas can be found in tau-equivalent latent

variable models, where facets loading onto a domain are exchangeable—that is, each personality characteristic is an equivalent manifestation of the overall domain (Edwards & Bagozzi, 2000; Schmittmann et al., 2013). Even without such a strict interpretation, there is still a pervasive idea that best practice is to design scales with uniform measurement (i.e., an equal number of items). This notion leads to unintended and potentially problematic consequences, particularly for sum scores (e.g., equal measurement of central and peripheral facets, not measuring what the researcher intends to measure, reduced or overly complex hierarchical structures). Therefore, the centrality of dimensions should be considered.

The notion of central and peripheral dimensions of a trait domain is not necessarily new (e.g., Connelly et al., 2014b; Roberts et al., 2005); however, their consideration in scale development is. Soto and John's (2017) BFI-2 is one of the first inventories to specify the centrality of their intended facets (one pure and two complementary per domain). The number of facets to include depends on the developer's intent: narrow inventories should emphasize central facets, while broad inventories should emphasize both central and peripheral facets. Regardless of intent, specific consideration should be given to the centrality of facets to increase each domain's test validity—that is, the test's sensitivity to variation in the psychological attribute (Borsboom, Cramer, Kievit, Scholten, & Franić, 2009).

Considering that most personality research uses sum scores when computing personality–outcome associations, more central facets should have greater substantive weight because they are considered "purer" assessments of the overall domain. This means that more central facets should have more items than more peripheral facets because central facets are more likely to have greater sensitivity to variation in the target domain than peripheral facets. Ultimately, this suggests that researchers *should* develop asymmetric scales (e.g., Watson et al., 2017). Furthermore, peripheral facets may contain items that are compound, meaning they measure another domain in a similar or greater degree. In a network, compound items would be depicted by having more or stronger inter-facet connections than intra-facet connections. Thus, if a researcher is attempting to measure a

pure domain, some items in peripheral facets can be removed (based on their connections to other facets) to increase domain and facet specificity (Mõttus, 2016).

The CCC and CEC are two measures that researchers can use to determine the centrality of their dimensions. Larger values would suggest that facets are more central while smaller values suggest that facets are more peripheral. Depending upon the scale developer's theoretical perspective, CCC would be recommended for a latent variable approach, while CEC would be recommended for a CTT approach. In either case, it may be useful for researchers to apply both measures to determine each dimension's centrality based on general centrality in the network (i.e., CCC) or centrality based on connectivity to other dimensions in the network (i.e., CEC).

**3.3.3. Item selection.** Traditional methods rely on unidimensional (CTT and IRT) or constrained multidimensional (MIRT and factor analysis) item selection analyses—that is, a simple structure, rather than complex, is used to select items. CTT and IRT, for example, have an assumption that scales are unidimensional and their selection criteria (item discrimination) is restricted to associations within the scale. MIRT and EFA do not have unidimensionality assumptions but only permit item-by-item criteria for each factor.[2] Despite many acknowledging that such a simple structure is merely a conceptual convenience (e.g., Costa & McCrae, 1995), nearly all current item selection practices (and by consequence nearly all existing scales) are based on simple structure approaches (Hubley, Zhu, Sasaki, & Gadermann, 2014). These approaches generally emphasize high item homogeneity, leading to scales that often reflect "bloated specifics" (Cattell, 1978), which is at odds with many modern personality theories and perspectives (McCrae & Mõttus, in press).

The network approach suggests that personality traits are composed of mutually reinforcing causal components (Cramer et al., 2012a). As a consequence of this perspective,

[2] Although EFA allows a complex structure, the item-by-item criteria (i.e., factor loadings) makes the identification of the item's relations to other items an onerous task. Furthermore, each item's relative location in multidimensional space is not immediately evident, resulting in decisions often made in the context of a single dimension. In contrast, psychometric networks make this information immediately apparent while also providing a more parsimonious model.

behavioral components represented by items should be relatively unique, rather than redundant, to reduce latent confounding (Hallquist et al., 2019). This stands in stark contrast to common scale development practices because traditional guidelines encourage redundancy to increase internal consistency. Thus, mostly unique items will inevitably lead to scales that have lower internal consistency. Lower internal consistency, however, should not be confused with lower validity or even lower reliability (e.g., Christensen et al., 2018b). In fact, scales with lower internal consistency tend to have higher validity (known as the reliability-validity paradox; Brennan, 2001), while demonstrating comparable test-retest reliability (McCrae, 2015). McCrae and Mõttus (in press) propose that scales should be composed of unique items that reflect a distinct nuance of a trait to maximize the scale's information and efficiency.

To identify unique items, a topological overlap measure can be used, which quantifies how similar a node's connections to another node's connections are (i.e., the common connections between any two nodes). In biological networks, measures of topological overlap have been used to identify genes or proteins that may have a similar biological pathway or function (Nowick, Gernat, Almaas, & Stubbs, 2009; Zhang & Horvath, 2005). Greater topological overlap suggests that two genes may belong to the same functional class compared to those with less overlap. Similarly, nodes that have high topological overlap in a personality network are likely to have a shared latent influence. By removing these nodes, unique components of the personality system can be uncovered, reducing the latent confounding between items.

With this in mind, items should not only be selected based on some item criteria but also by their connections to other items in multidimensional space. The main vehicle for item selection criteria in psychometric networks is node strength because it offers the most direct equivalence to factor loadings and item-scale correlations (Hallquist et al., 2019). A notable consideration is that a node's strength can potentially be a blend of connections within and between dimensions, which may not reflect a strong unidimensional indicator (Hallquist et al.,

2019). Researchers should consider each item's within- and between-dimension node strength when selecting their items. In short, node strength represents the best singular item selection criterion; however, it should be tempered within the context of dimensional relations.

Christensen and Golino's (2019) bootEGA approach offers two metrics—item stability and item statistics—that directly address this issue. The item stability metric quantifies the number of times an item is replicated in a researcher specified dimension (e.g., exploratory graph analysis). Lower item replication values in the specified dimension might suggest item misallocation (e.g., $\leq$.25), while equal replication values between two or more dimensions might suggest multidimensionality. The item statistics metric takes advantage of the intra- and inter-dimension node strength measures (e.g., Blanken et al., 2018) to determine how strong the connections for each item are within each dimension. In this context, the item statistics metric appears to be equivalent to EFA factor loadings. When these metrics are used in combination, researchers can identify whether certain items are compound, misspecified, or problematic for the dimensional structure of the scale.

Other centrality measures potentially offer additional criteria for item selection. Nodes that are high in betweenness, for example, have been referred to as "bridges" between two dimensions in the network (Cramer et al., 2010). In Hallquist and colleagues' (2019) simulations, however, they found that betweenness was susceptible to sampling variation and may reflect spurious associations between items in different latent dimensions rather than substantive bridges. One potential solution to overcome this issue would be to use a measure called the *randomized shortest path betweenness centrality* (RSPBC; Kivimäki, Lebichot, Saramäki, & Saerens, 2016). The RSPBC remedies some limitations of the standard betweenness centrality by using random walks—random steps from one node to the next—over the network. By using random walks, rather than the relative shortest path between nodes, there is greater discrimination of nodes that are clearly *between* other nodes (Christensen, 2018). Although there has yet to be an investigation into the stability of the RSPBC, it's likely that it will be less susceptible to dramatic shifts in network connectivity

than the standard betweenness centrality measure.

The RSPBC offers a complementary criterion to node strength for identifying nodes that may be bridges or multidimensional. Nodes that are high RSPBC but have low inter-dimension node strength, for example, would suggest that a node is a bridge. Conversely, nodes that are high RSPBC and have high inter-dimension node strength would suggest that a node is multidimensional. Christensen and Golino's (2019) item statistics metric could also be used to corroborate these classifications. In the context of item selection, a scale developer could consider removing or adding more bridge or multidimensional items to adjust the relations between dimensions, making them more or less orthogonal. Therefore, the combination of these measures offers more control over the theoretical attribute being measured, allowing the researcher to directly influence the dimensional structure of the scale.

Finally, the utility of the hybrid centrality still requires greater attention. Based on the simulation performance, the hybrid centrality is not as compelling as node strength in terms of item selection criteria. In general, items high in hybrid centrality could be used to determine which items are the most central to the overall latent construct. Indeed, these items are not only strongly connected to other items (high strength and eigenvector centrality) but are also more central within and between dimensions (high betweenness and closeness centrality). For this reason, the hybrid centrality could be used as an indicator for items that are suitable for developing abbreviated scales (Christensen et al., 2018b). It could also be used to test whether the dimensions are measuring the theoretical terms they are intended to measure (e.g., dimension labels; Christensen et al., 2018a), particularly in relation to the global latent construct.

**3.3.4. Summary.** These guidelines build on traditional scale development practices and insert new perspectives that are provided by psychometric networks. In general, psychometric networks provide several assessment advantages that other psychometric approaches do not. These advantages are primarily afforded by the novel representation of network models, which depict items in a complex structure (i.e., multidimensional space)

rather than examining them in a simple structure (i.e., isolated space). Topological overlap measures allow researchers to identify redundant nodes, reducing latent confounding between items. Node centrality measures offer similar information to traditional item selection measures (e.g., intra-dimension strength) but also whether certain items may belong to multiple latent dimensions (e.g., inter-dimension strength, RSPBC, bootEGA's metrics). In addition, the hybrid centrality may be an efficient measure for developing abbreviated versions of a scale. Community centrality measures allow developers to test whether their facets are as central or peripheral as they intend, leading to more accurate scale specifications. These measures also promote the development asymmetric scales, which should increase the scale's test validity. Finally, exploratory graph analysis offers the most accurate dimension identification approach to date, ensuring researchers achieve optimal and generalizable dimensionality results.

### 3.4. Areas for Future Research

**3.4.1. Reproducibility.** There are several key areas that should be developed further to establish psychometric networks in common scale development practices. The most recurrent issue in the literature is reproducibility (e.g., Borsboom, Robinaugh, Group, Rhemtulla, & Cramer, 2018; Christensen et al., 2018c; Forbes et al., 2017; Fried et al., 2018). This issue is not unique to networks as SEM and latent variable models have had their own growing pains in this area. The simple solution is larger samples; however, time and resources are not always available. Recent methodological developments using bootstrap approaches to quantify the stability of network measures provide a way for researchers to check whether their results would generalize to different or smaller samples (e.g., Christensen & Golino, 2019; Epskamp et al., 2018a). Notably, these approaches are starting to become standard practice in the psychometric network literature, which should only increase the likelihood of reproducible results.

**3.4.2. Generalizability.** Beyond bootstrap approaches, cross-validation techniques may provide more robust support for generalizability (Kan et al., 2019; Yarkoni & Westfall, 2017). Split-half sampling or multiple samples could be used to first fit a network structure to a test sample and then fit the same structure (i.e., identical edges) to a validation sample (Forbes et al., 2017). For a confirmatory approach, two independent samples could be obtained with one serving as the testing model and the other as the confirmatory model, which is then fit using SEM procedures (e.g., Kan et al., 2019). Traditional fit measures could be reported along with relative measures like Kullback-Leibler divergence and root mean square error (Williams & Rast, 2018). More specific to dimensional structure, similar approaches could be applied with the use of normalized mutual information, which could be computed between the exploratory graph analysis solutions to determine relative fit of dimensions (Danon, Diaz-Guilera, Duch, & Arenas, 2005).

**3.4.3. Unidimensionality.** One critical future direction for exploratory graph analysis is to examine its ability to detect the unidimensionality of scales. Because unidimensionality is a integral feature of standard psychometric practice, this warrants considerable attention (DeVellis, 2017; Flake, Pek, & Hehman, 2017; Simms, 2008). One obstacle that prevents current implementations of exploratory graph analysis from detecting one dimension solutions is community detection algorithms. These algorithms are designed to detect partitions in networks (particularly large networks; i.e., nodes $> 1000$) and almost always find partitions (i.e., 2 or more dimensions). Notably, however, the modularity statistic (Q), or how well the network is partitioned, could be used to decide whether a network is unidimensional. If two dimensions are returned, then the Q value should be evaluated. If Q is large, then it can be reasonably inferred that there are two dimensions. If Q is small, then this may suggest that there is only one dimension. Inevitably, simulation studies will be necessary to determine the cut-offs for what constitutes “large” and “small” Q values.

**3.4.4. Connecting within-person processes with between-person structure.** Perhaps the most important issue for future research is to establish how intraindividual

processes lead to interindividual structure (Baumert et al., 2017; Beck & Jackson, 2017; Borkenau & Ostendorf, 1998; Borsboom et al., 2009; Breil et al., 2019). A fundamental assumption of the network approach is that the variables in interindividual networks are causally coupled relations, which represent underlying processes (e.g., Borsboom, 2017; Cramer et al., 2012a); however, it still needs to be tested whether this is actually the case. Epskamp and colleagues' (2018c) MNM framework provides a vehicle to test this interpretation. One method would be to administer a personality questionnaire at two disparate time points (e.g., a month apart) and administer EMA questions that reflect items in the questionnaire. If the interindividual stability (and variability) of edges in the questionnaire's network corroborate the EMA measurement, then it can be reasonably inferred that they represent the underlying processes. Other methods include using daily diaries (Zimmermann et al., in press) as well as Electronic Activated Recorders to unobtrusively capture audio in naturalistic environments (Sun & Vazire, 2019). In general, EMA approaches combined with the MNM framework are pivotal for the integration of intraindividual processes with interindividual structure (Baumert et al., 2017; Beck & Jackson, 2017).

**3.4.5. Interpreation and psychometric properties of centrality measures.** The MNM framework could also help resolve how centrality measures should be interpreted (Bringmann et al., 2018). Although the small-scale simulation presented in this paper took one step towards determining how certain centrality measures are associated with more traditional psychometric measures, their interpretations in terms of processes are still unclear. If centrality of interindividual nodes have meaning that is connected to intraindividual processes, then their interpretations could be clarified and validated. In general, we echo the call of Hallquist and colleagues (2019) for more studies using simulated and real-world data to further flesh out the meaning of centrality measures and their significance in psychometric theory. Future research, for example, should attempt to elucidate the stability of RSPBC and hybrid centrality and determine their role (if any) for evaluating the psychometric

properties of scales. Other directions include attempting to see how measures, such as topological overlap, can discern the redundancy of items, so that unique behavioral components can be more concretely identified. The importance of methods that can detect and increase the uniqueness of items in personality scales extends beyond networks but to personality psychometrics more generally (McCrae & Mõttus, in press).

## 4. Concluding Remarks

The structure of personality traits is complex, yet it has not always been measured and assessed this way. On the one hand, there has been a lack of integration of personality theory into scale development practices. On the other hand, there has been a reluctance to adopt more sophisticated psychometric perspectives. With the rise of network psychometrics, a new wave of measurement perspectives has emerged, providing an opportunity to refine measurement and assessment practices. In this paper, we've elaborated on how psychometric networks go beyond a novel measurement perspective by demonstrating how they can inform personality theory and provide a representation of interactions between variables that aligns well with modern measurement trends in personality research. From this, we advanced a nascent framework for developing personality assessment instruments that represent personality as a complex interconnected network of behaviors. With this new tool in hand, researchers can begin to build scales that reflect the full complexity of personality, paving the way for new advances in personality theory and measurement.

## References

Allport, G. W. (1937). *Personality: A psychological interpretation*. Oxford, UK: Holt.

Anglim, J., & Grant, S. (2016). Predicting psychological and subjective well-being from personality: Incremental prediction from 30 facets over the Big 5. *Journal of Happiness Studies*, *17*, 59–80. https://doi.org/10.1007/s10902-014-9583-7

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 397–438. https://doi.org/10.1080/10705510903008204

Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P., ... Wrzus, C. (2017). Integrating personality structure, personality process, and personality development. *European Journal of Personality*, *31*, 503–528. https://doi.org/10.1002/per.2115

Beck, E. D., & Jackson, J. J. (2017). The search for a bridge: Idiographic personality networks. *European Journal of Personality*, *31*, 530–532. https://doi.org/10.1002/per

Bell, V., & O'Driscoll, C. (2018). The network structure of paranoia in the general population. *Social Psychiatry and Psychiatric Epidemiology*, *53*, 1–8. https://doi.org/10.1007/s00127-018-1487-0

Blain, S. D., Longenecker, J. M., Grazioplene, R. G., Klimes-Dougan, B., & DeYoung, C. G. (2019). Apophenia as the disposition to false positives: A unifying framework for positive symptoms across the openness-psychoticism spectrum. *PsyArXiv*. https://doi.org/10.31234/osf.io/d9wkc

Blanken, T. F., Deserno, M. K., Dalege, J., Borsboom, D., Blanken, P., Kerkhof, G. A., & Cramer, A. O. (2018). The role of stabilizing and communicating symptoms given overlapping communities in psychopathology networks. *Scientific Reports*, *8*, 5854. https://doi.org/10.1038/s41598-018-24224-2

Bleidorn, W., Hopwood, C. J., Ackerman, R. A., Witt, E. A., Kandler, C., Riemann, R., ... Donnellan, M. B. (in press). The healthy personality from a basic trait perspective.

*Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspp0000231

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*, 187–215. https://doi.org/10.1037/0033-2909.117.2.187

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*, 605–634. https://doi.org/10.1146/annurev.psych.53.100901.135239

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*, 113–120. https://doi.org/10.1080/0022250X.1972.9989806

Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor model to describe intraindividual variations over time? *Journal of Research in Personality*, *32*, 202–221. https://doi.org/10.1006/jrpe.1997.2206

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*, 425–440. https://doi.org/10.1007/s11336-006-1447-6

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

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

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

Borsboom, D., Cramer, A. O., Kievit, R. A., Scholten, A. Z., & Franić, S. (2009). The end of construct validity. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 135–170). Charlotte, NC: IAP Information Age Publishing.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219.

https://doi.org/10.1037/0033-295X.110.2.203

Borsboom, D., Robinaugh, D. J., Group, P., Rhemtulla, M., & Cramer, A. O. (2018). Robustness and replicability of psychopathology networks. *World Psychiatry*, *17*, 143–144. https://doi.org/10.1002/wps.20515

Breil, S. M., Geukes, K., Wilson, R. E., Nestler, S., Vazire, S., & Back, M. D. (2019). Zooming into real-life extraversion–How personality and situation shape sociability in social interactions. *Collabra: Psychology*, *5*, 7. https://doi.org/10.1525/collabra.170

Brennan, R. L. (2001). Some problems, pitfalls, and paradoxes in educational measurement. *Educational Measurement: Issues and Practice*, *20*, 6–18. https://doi.org/10.1111/j.1745-3992.2001.tb00071.x

Bringmann, L. F., & Eronen, M. I. (2018). Don't blame the model: Reconsidering the network approach to psychopathology. *Psychological Review*, *125*, 606–615. https://doi.org/10.1037/rev0000108

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., . . . Snippe, E. (2018). What do centrality measures measure in psychology networks? Retrieved from: https://doi.org/10.13140/RG.2.2.25024.58884

Carter, N. T., Miller, J. D., & Widiger, T. A. (2018). Extreme personalities at work and in life. *Current Directions in Psychological Science*, *27*, 429–436. https://doi.org/10.1177/0963721418793134

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. Boston, MA: Springer. https://doi.org/10.1007/978-1-4684-2262-7

Cervone, D. (2005). Personality architecture: Within-person structures and processes. *Annual Review of Psychology*, *56*, 423–452. https://doi.org/10.1146/annurev.psych.56.091103.070133

Chernyshenko, O. S., Stark, S., Drasgow, F., & Roberts, B. W. (2007). Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment*, *19*,

88–106. https://doi.org/10.1037/1040-3590.19.1.88

Christensen, A. P. (2018). NetworkToolbox: Methods and measures for brain, cognitive, and psychometric network analysis in R. *The R Journal*, *10*, 422–439. https://doi.org/10.32614/RJ-2018-065

Christensen, A. P., & Golino, H. F. (2019). Estimating the stability of the number of factors via Bootstrap Exploratory Graph Analysis: A tutorial. *PsyArXiv*. https://doi.org/10.31234/osf.io/9deay

Christensen, A. P., Cotter, K. N., & Silvia, P. J. (2018a). Reopening openness to experience: A network analysis of four openness to experience inventories. *Journal of Personality Assessment*. https://doi.org/10.1080/00223891.2018.1467428

Christensen, A. P., Cotter, K., Silvia, P., & Benedek, M. (2018b). Scale development via network analysis: A comprehensive and concise measure of openness to experience. *PsyArXiv*. https://doi.org/10.31234/osf.io/3raxt

Christensen, A. P., Kenett, Y. N., Aste, T., Silvia, P. J., & Kwapil, T. R. (2018c). Network structure of the Wisconsin Schizotypy Scales–Short Forms: Examining psychometric network filtering approaches. *Behavior Research Methods*, *50*, 2531–2550. https://doi.org/10.3758/s13428-018-1032-9

Christensen, A. P., Kenett, Y. N., Cotter, K. N., Beaty, R. E., & Silvia, P. J. (2018). Remotely close associations: Openness to experience and semantic memory structure. *European Journal of Personality*, *32*, 480–492. https://doi.org/10.1002/per.2157

Clark, L. A., & Livesley, W. J. (2002). Two approaches to identifying the dimensions of personality disorder: Convergence on the five-factor model. In P. T. Costa & T. A. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (pp. 161–176). Washington, DC: American Psychological Association. https://doi.org/10.1037/10423-010

Connelly, B. S., Ones, D. S., & Chernyshenko, O. S. (2014a). Introducing the special section on openness to experience: Review of openness taxonomies, measurement, and

nomological net. *Journal of Personality Assessment*, *96*, 1–16. https://doi.org/10.1080/00223891.2013.830620

Connelly, B. S., Ones, D. S., Davies, S. E., & Birkland, A. (2014b). Opening up openness: A theoretical sort following critical incidents methodology and a meta-analytic investigation of the trait family measures. *Journal of Personality Assessment*, *96*, 17–28. https://doi.org/10.1080/00223891.2013.809355

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, *64*, 21–50. https://doi.org/10.1207/s15327752jpa6401_2

Costantini, G., & Perugini, M. (2016). The network of conscientiousness. *Journal of Research in Personality*, *65*, 68–88. https://doi.org/10.1016/j.jrp.2016.10.003

Costantini, G., & Perugini, M. (in press). Network analysis for psychological situations. In J. F. Rauthmann, R. Sherman, & D. C. Funder (Eds.), *The Oxford handbook of psychological situations* (pp. 1–41). Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263348.013.16

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. (2015a). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality*, *54*, 13–29. https://doi.org/10.1016/j.jrp.2014.07.003

Costantini, G., Richetin, J., Borsboom, D., Fried, E. I., Rhemtulla, M., & Perugini, M. (2015b). Development of indirect measures of conscientiousness: Combining a facets approach and network analysis. *European Journal of Personality*, *29*, 548–567. https://doi.org/10.1002/per.2014

Costantini, G., Richetin, J., Preti, E., Casini, E., Epskamp, S., & Perugini, M. (2019).

Stability and variability of personality networks. A tutorial on recent developments in network psychometrics. *Personality and Individual Differences*, *136*, 68–78. https://doi.org/10.1016/j.paid.2017.06.011

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

Cramer, A. O., Van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., ... Borsboom, D. (2012b). Measurable like temperature or mereological like flocking? On the nature of personality traits. *European Journal of Personality*, *26*, 451–459. https://doi.org/10.1002/per.1879

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

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

Danaher, P., Wang, P., & Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*, 373–397. https://doi.org/10.1111/rssb.12033

Danon, L., Diaz-Guilera, A., Duch, J., & Arenas, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, *2005*, P09008. https://doi.org/10.1088/1742-5468/2005/09/P09008

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Thousand Oaks, CA: SAGE Publications.

DeYoung, C. G. (2015). Cybernetic Big Five theory. *Journal of Research in Personality*, *56*,

33–58. https://doi.org/10.1016/j.jrp.2014.07.004

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, *33*, 533–552. https://doi.org/10.1016/S0191-8869(01)00171-4

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*, 880–896. https://doi.org/10.1037/0022-3514.93.5.880

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, *73*, 1246–1256. https://doi.org/10.1037/0022-3514.73.6.1246

Dunlop, W. L. (2015). Contextualized personality, beyond traits. *European Journal of Personality*, *29*, 310–325. https://doi.org/10.1002/per.1995

Durbin, C. E., & Hicks, B. M. (2014). Personality and psychopathology: A stagnant field in need of development. *European Journal of Personality*, *28*, 362–386. https://doi.org/10.1002/per.1962

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

Ellis, J. L., & van den Wollenberg, A. L. (1993). Local homogeneity in latent trait models: A characterization of the homogeneous monotone IRT model. *Psychometrika*, *58*, 417–429. https://doi.org/10.1007/BF02294649

Embretson, S. E. (2004). The second century of ability testing: Some predictions and speculations. *Measurement: Interdisciplinary Research and Perspectives*, *2*, 1–32. https://doi.org/10.1207/s15366359mea0201_1

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*, 617–634. https://doi.org/10.1037/met0000167

Epskamp, S., Borsboom, D., & Fried, E. I. (2018a). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*, 195–212.

https://doi.org/10.3758/s13428-017-0862-1

Epskamp, S., Maris, G., Waldrop, L. J., & Borsboom, D. (2018b). Network psychometrics. In P. Irwing, D. Hughes, & T. Booth (Eds.), *The Wiley handbook of psychometric testing, 2 volume set: A multidisciplinary reference on survey, scale and test development*. New York, NY: Wiley. https://doi.org/10.1002/9781118489772.ch30

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

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018c). The Gaussian Graphical Model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *4*, 1–28. https://doi.org/10.1080/00273171.2018.1454823

Everett, M. G., & Borgatti, S. P. (1999). The centrality of groups and classes. *The Journal of Mathematical Sociology*, *23*, 181–201. https://doi.org/10.1080/0022250X.1999.9990219

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, *8*, 370–378. https://doi.org/10.1177/1948550617693063

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*, 1011–1027. https://doi.org/10.1037/0022-3514.80.6.1011

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, *13*, 83–87. https://doi.org/10.1111/j.0963-7214.2004.00280.x

Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. *Journal of Research in Personality*, *56*, 82–92. https://doi.org/10.1016/j.jrp.2014.10.009

Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale

development and validation. *Canadian Journal of Behavioural Science*, *49*, 78–88. https://doi.org/10.1037/cbs0000069

Forbes, M. K., Wright, A. G., Markon, K. E., & Krueger, R. F. (2017). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology*, *126*, 969–988. https://doi.org/10.1037/abn0000276

Fried, E. I., Eidhof, M. B., Palic, S., Costantini, G., Huisman-van Dijk, H. M., Bockting, C. L., . . . Karstoft, K.-I. (2018). Replicability and generalizability of posttraumatic stress disorder (PTSD) networks: A cross-cultural multisite study of PTSD symptoms in four trauma patient samples. *Clinical Psychological Science*, *6*, 335–351. https://doi.org/10.1177/2167702617745092

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*, 432–441. https://doi.org/10.1093/biostatistics/kxm045

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2011). Performance of Velicer's minimum average partial factor retention method with categorical variables. *Educational and Psychological Measurement*, *71*, 551–570. https://doi.org/10.1177/0013164410389489

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods*, *18*, 454–474. https://doi.org/10.1037/a0030005

Giscard, P.-L., & Wilson, R. C. (2018). A centrality measure for cycles and subgraphs II. *Applied Network Science*, *3*, 1–15. https://doi.org/10.1007/s41109-018-0064-5

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*, 26–34. https://doi.org/10.1037/0003-066X.48.1.26

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28).

Tilburg, The Netherlands: Tilburg University Press.

Golino, H. F., & Demetriou, A. (2017). Estimating the dimensionality of intelligence like data using Exploratory Graph Analysis. *Intelligence*, *62*, 54–70. https://doi.org/10.1016/j.intell.2017.02.007

Golino, H. F., & Epskamp, S. (2017). Exploratory Graph Analysis: A new approach for estimating the number of dimensions in psychological research. *PloS ONE*, *12*, e0174035. https://doi.org/10.1371/journal.pone.0174035

Golino, H., Shi, D., Garrido, L. E., Christensen, A. P., Nieto, M. D., Sadana, R., & Thiyagarajan, J. A. (2018). Investigating the performance of Exploratory Graph Analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *PsyArXiv*. https://doi.org/10.31234/osf.io/gzcre

Hallquist, M., Wright, A. C. G., & Molenaar, P. C. M. (2019). Problems with centrality measures in psychopathology symptom networks: Why network psychometrics cannot escape psychometric theory. *PsyArXiv*. https://doi.org/10.31234/osf.io/pg4mf

Hamaker, E. L., Nesselroade, J. R., & Molenaar, P. C. M. (2007). The integrated trait–state model. *Journal of Research in Personality*, *41*, 295–315. https://doi.org/10.1016/j.jrp.2006.04.003

Haslbeck, J. M. B., & Waldorp, L. J. (2018). How well do network models predict observations? On the importance of predictability in network models. *Behavior Research Methods*, *50*, 853–861. https://doi.org/10.3758/s13428-017-0910-x

Herrmann, A., & Pfister, H.-R. (2013). Simple measures and complex structures: Is it worth employing a more complex model of personality in Big Five inventories? *Journal of Research in Personality*, *47*, 599–608. https://doi.org/10.1016/j.jrp.2013.05.004

Hubley, A. M., Zhu, S. M., Sasaki, A., & Gadermann, A. M. (2014). Synthesis of validation practices in two assessment journals: Psychological Assessment and the European Journal of Psychological Assessment. In B. Zumbo & E. Chan (Eds.), *Validity and validation in social, behavioral, and health sciences* (pp. 193–213). Cham, CH:

Springer.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 159–181). New York, NY: Guilford Press.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory—Versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

Kan, K.-J., van der Maas, H. L., & Levine, S. Z. (2019). Extending psychometric network analysis: Empirical evidence against $g$ in favor of mutualism? *Intelligence*, *73*, 52–62. https://doi.org/10.1016/j.intell.2018.12.004

Kaufman, S. B., Quilty, L. C., Grazioplene, R. G., Hirsh, J. B., Gray, J. R., Peterson, J. B., & DeYoung, C. G. (2016). Openness to experience and intellect differentially predict creative achievement in the arts and sciences. *Journal of Personality*, *84*, 248–258. https://doi.org/10.1111/jopy.12156

Kievit, R. A., Van Rooijen, H., Wicherts, J. M., Waldorp, L. J., Kan, K.-J., Scholte, H. S., & Borsboom, D. (2012). Intelligence and the brain: A model-based approach. *Cognitive Neuroscience*, *3*, 89–97. https://doi.org/10.1080/17588928.2011.628383

Kivimäki, I., Lebichot, B., Saramäki, J., & Saerens, M. (2016). Two betweenness centrality measures based on randomized shortest paths. *Scientific Reports*, *6*, 19668. https://doi.org/10.1038/srep19668

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine*, *42*, 1879–1890. https://doi.org/10.1017/S0033291711002674

Kruis, J., & Maris, G. (2016). Three representations of the Ising model. *Scientific Reports*, *6*,

srep34175. https://doi.org/10.1038/srep34175

Lauritzen, S. L. (1996). *Graphical models*. Oxford, UK: Clarendon Press.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology*, *88*, 139–157. https://doi.org/10.1037/0022-3514.88.1.139

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L., . . . Maris, G. (2018). An introduction to network psychometrics: Relating Ising network models to item response theory models. *Multivariate Behavioral Research*, *53*, 15–35. https://doi.org/10.1080/00273171.2017.1379379

Massara, G. P., Di Matteo, T., & Aste, T. (2016). Network filtering for big data: Triangulated Maximally Filtered Graph. *Journal of Complex Networks*, *5*, 161–178. https://doi.org/10.1093/comnet/cnw015

McAdams, D. P. (1992). The five-factor model in personality: A critical appraisal. *Journal of Personality*, *60*, 329–361. https://doi.org/10.1111/j.1467-6494.1992.tb00976.x

McCabe, K. O., & Fleeson, W. (2012). What is extraversion for? Integrating trait and motivational perspectives and identifying the purpose of extraversion. *Psychological Science*, *23*, 1498–1505. https://doi.org/10.1177/0956797612444904

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review*, *19*, 97–112. https://doi.org/10.1177/1088868314541857

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*, 81–90. https://doi.org/10.1037/0022-3514.52.1.81

McCrae, R. R., & Costa, P. T. (1996). Toward a new generation of personality theories:

Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 51–87). New York, NY: Guilford Press.

McCrae, R. R., & Costa, P. T. (2008). The five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 159–181). New York, NY: Guilford Press.

McCrae, R. R., & Mõttus, R. (in press). A new psychometrics: What personality scales measure, with implications for theory and assessment. *Current Directions in Psychological Science*.

McCrae, R. R., Zonderman, A. B., Costa, P. T., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, *70*, 552–566. https://doi.org/10.1037/0022-3514.70.3.552

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246–268. https://doi.org/10.1037/0033-295X.102.2.246

Mischel, W., & Shoda, Y. (2008). Toward a unified theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 208–241). New York, NY: Guilford Press.

Molenaar, P. C. M. (2010). Latent variable models are network models. *Behavioral and Brain Sciences*, *33*, 166–166. https://doi.org/10.1017/S0140525X10000798

Molenaar, P. C. M., van Rijn, P., & Hamaker, E. (2007). A new class of SEM model equivalences and its implications. In S. M. Boker & M. J. Wenger (Eds.), *Data analytic techniques for dynamical systems* (pp. 189–211). Mahwah, NJ: Erlbaum.

Mõttus, R. (2016). Towards more rigorous personality trait–outcome research. *European*

*Journal of Personality*, *30*, 292–303. https://doi.org/10.1002/per.2041

Mõttus, R., & Allerhand, M. (2017). Why do traits come together? The underlying trait and network approaches. In V. Ziegler-Hill & T. K. Shackelford (Eds.), *SAGE handbook of personality and individual differences: The science of personality and individual differences* (pp. 1–22). London, UK: SAGE Publications.

Mõttus, R., Allerhand, M., & Johnson, W. (in pressa). Computational modeling of person-situation accumulation of situational experience can shape the distribution of trait scores. In J. F. Rauthmann, R. Sherman, & D. C. Funder (Eds.), *The Oxford handbook of psychological situations* (pp. 1–41). Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263348.013.23

Mõttus, R., Bates, T., Condon, D. M., Mroczek, D., & Revelle, W. (2018). Your personality data can do more: Items provide leverage for explaining the variance and co-variance of life outcomes. *PsyArXiv*. https://doi.org/10.31234/osf.io/4q9gv

Mõttus, R., Epskamp, S., & Francis, A. (2017). Within- and between-individual variability of personality characteristics and physical exercise. *Journal of Research in Personality*, *69*, 139–148. https://doi.org/10.1016/j.jrp.2016.06.017

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, *112*, 474–490. https://doi.org/10.1037/pspp0000100

Mõttus, R., Sinick, J., Terracciano, A., Hřebíčková, M., Kandler, C., Ando, J., . . . Jang, K. L. (in pressb). Personality characteristics below facets: A replication and meta-analysis of cross-rater agreement, rank-order stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspp0000202

Nowick, K., Gernat, T., Almaas, E., & Stubbs, L. (2009). Differences in human and chimpanzee gene expression patterns define an evolving network of transcription

factors in brain. *Proceedings of the National Academy of Sciences*, *106*, 22358–22363. https://doi.org/10.1073/pnas.0911376106

Perugini, M., Costantini, G., Hughes, S., & De Houwer, J. (2016). A functional perspective on personality. *International Journal of Psychology*, *51*, 33–39. https://doi.org/10.1002/ijop.12175

Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, *10*, 191–218. https://doi.org/10.7155/jgaa.00185

Pozzi, F., Di Matteo, T., & Aste, T. (2013). Spread of risk across financial markets: Better to invest in the peripheries. *Scientific Reports*, *3*, 1665. https://doi.org/10.1038/srep01665

Read, S. J., & Miller, L. C. (2002). Virtual personalities: A neural network model of personality. *Personality and Social Psychology Review*, *6*, 357–369. https://doi.org/10.1207/S15327957PSPR0604_10

Read, S. J., Droutman, V., & Miller, L. C. (2017). Virtual personalities: A neural network model of the structure and dynamics of personality. In R. R. Vallacher, S. J. Read, & A. Nowak (Eds.), *Computational social psychology* (pp. 31–53). New York, NY: Routledge.

Read, S. J., Monroe, B. M., Brownstein, A. L., Yang, Y., Chopra, G., & Miller, L. C. (2010). A neural network model of the structure and dynamics of human personality. *Psychological Review*, *117*, 61–92. https://doi.org/10.1037/a0018131

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, *12*, 287–297. https://doi.org/10.1037/1040-3590.12.3.287

Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology*, *58*, 103–139.

https://doi.org/10.1111/j.1744-6570.2005.00301.x

Samuel, D. B., & Widiger, T. A. (2008). A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: A facet level analysis. *Clinical Psychology Review*, *28*, 1326–1342. https://doi.org/10.1016/j.cpr.2008.07.002

Samuel, D. B., Simms, L. J., Clark, L. A., Livesley, W. J., & Widiger, T. A. (2010). An item response theory integration of normal and abnormal personality scales. *Personality Disorders: Theory, Research, and Treatment*, *1*, 5–21. https://doi.org/10.1037/a0018136

Saucier, G., & Goldberg, L. R. (1996). The language of personality: Lexical perspectives on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality* (pp. 21–50). New York, NY: Guilford Press.

Saulsman, L. M., & Page, A. C. (2004). The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clinical Psychology Review*, *23*, 1055–1085. https://doi.org/10.1016/j.cpr.2002.09.001

Schmittmann, V. D., Cramer, A. O., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*, *31*, 43–53. https://doi.org/10.1016/j.newideapsych.2011.02.007

Seeboth, A., & Mõttus, R. (2018). Successful explanations start with accurate descriptions: Questionnaire items as personality markers for more accurate predictions. *European Journal of Personality*, *32*, 186–201. https://doi.org/10.1002/per.2147

Shoda, Y., LeeTiernan, S., & Mischel, W. (2002). Personality as a dynamical system: Emergence of stability and distinctiveness from intra and interpersonal interactions. *Personality and Social Psychology Review*, *6*, 316–325. https://doi.org/10.1207/S15327957PSPR0604_06

Simms, E. E. (2009). *Assessment of the facets of the five-factor model: Further development and validation of a new personality measures* (Unpublished doctoral dissertation).

University of Iowa, Iowa City, IA, USA.

Simms, L. J. (2008). Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass*, *2*, 414–433. https://doi.org/10.1111/j.1751-9004.2007.00044.x

Sosnowska, J., Kuppens, P., De Fruyt, F., & Hofmans, J. (2019). A dynamic systems approach to personality: The Personality Dynamics (PersDyn) model. *Personality and Individual Differences*, *144*, 11–18. https://doi.org/10.1016/j.paid.2019.02.013

Soto, C. J., & John, O. P. (2009). Ten facet scales for the Big Five Inventory: Convergence with NEO PI-R facets, self-peer agreement, and discriminant validity. *Journal of Research in Personality*, *43*, 84–90. https://doi.org/10.1016/j.jrp.2008.10.002

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*, 117–143. https://doi.org/10.1037/pspp0000096

Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, *91*, 25–39. https://doi.org/10.1037/0021-9010.91.1.25

Stepp, S. D., Yu, L., Miller, J. D., Hallquist, M. N., Trull, T. J., & Pilkonis, P. A. (2012). Integrating competing dimensional models of personality: Linking the SNAP, TCI, and NEO using item response theory. *Personality Disorders: Theory, Research, and Treatment*, *3*, 107–126. https://doi.org/10.1037/a0025905

Sun, J., & Vazire, S. (2019). Do people know what they're like in the moment? *Psychological Science*. https://doi.org/10.1177/0956797618818476

Tay, L., & Jebb, A. T. (2018). Establishing construct continua in construct validation: The process of continuum specification. *Advances in Methods and Practices in*

*Psychological Science*, *1*, 375–388. https://doi.org/10.1177/2515245918775707

Thomas, K. M., Yalch, M. M., Krueger, R. F., Wright, A. G., Markon, K. E., & Hopwood, C. J. (2013). The convergent structure of DSM-5 personality trait facets and five-factor model trait domains. *Assessment*, *20*, 308–311. https://doi.org/10.1177/1073191112457589

Thurstone, L. L. (1947). *Multiple-factor analysis; a development and expansion of The Vectors of Mind*. Chicago, IL: University of Chicago Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288. Retrieved from jstor.org/stable/2346178

van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports*, *4*, 5918. https://doi.org/10.1038/srep05918

van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842–861. https://doi.org/10.1037/0033-295X.113.4.842

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, *90*, 7–24. https://doi.org/10.1016/j.ajhg.2011.11.029

Waller, N. G., DeYoung, C. G., & Bouchard, T. J. (2016). The recaptured scale technique: A method for testing the structural robustness of personality scales. *Multivariate Behavioral Research*, *51*, 433–445. https://doi.org/10.1080/00273171.2016.1157753

Walton, K. E., Roberts, B. W., Krueger, R. F., Blonigen, D. M., & Hicks, B. M. (2008). Capturing abnormal personality with normal personality inventories: An item response theory approach. *Journal of Personality*, *76*, 1623–1648.

https://doi.org/10.1111/j.1467-6494.2008.00533.x

Watson, D., Clark, L. A., & Chmielewski, M. (2008). Structures of personality and their relevance to psychopathology: II. Further articulation of a comprehensive unified trait structure. *Journal of Personality*, *76*, 1545–1586. https://doi.org/10.1111/j.1467-6494.2008.00531.x

Watson, D., Ellickson-Larew, S., Stanton, K., & Levin-Aspenson, H. (2016). Personality provides a general structural framework for psychopathology: Commentary on "Translational applications of personality science for the conceptualization and treatment of psychopathology". *Clinical Psychology: Science and Practice*, *23*, 309–313. https://doi.org/10.1111/cpsp.12164

Watson, D., Nus, E., & Wu, K. D. (2017). Development and validation of the faceted inventory of the five-factor model (FI-FFM). *Assessment*, 1073191117711022. https://doi.org/10.1177/1073191117711022

Watson, D., Stasik, S. M., Ellickson-Larew, S., & Stanton, K. (2015). Extraversion and psychopathology: A facet-level analysis. *Journal of Abnormal Psychology*, *124*, 432–446. https://doi.org/10.1037/abn0000051

Weiss, A., Gale, C. R., Batty, G. D., & Deary, I. J. (2013). A questionnaire-wide association study of personality and mortality: The Vietnam Experience Study. *Journal of Psychosomatic Research*, *74*, 523–529. https://doi.org/10.1016/j.jpsychores.2013.02.010

Widiger, T. A., & Crego, C. (in press). The bipolarity of normal and abnormal personality structure: Implications for assessment. *Psychological Assessment*.

Widiger, T. A., & Simonsen, E. (2005). Alternative dimensional models of personality disorder: Finding a common ground. *Journal of Personality Disorders*, *19*, 110–130. https://doi.org/10.1521/pedi.19.2.110.62628

Widiger, T. A., Gore, W. L., Crego, C., Rojas, S. L., & Oltmanns, J. (2017). Five factor model and personality disorder. In T. A. Widiger (Ed.), *The Oxford handbook of the*

*five-factor model* (pp. 449–478). New York, NY: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199352487.013.4

Widiger, T. A., Sellbom, M., Chmielewski, M., Clark, L. A., DeYoung, C. G., Kotov, R., ... Wright, A. G. C. (2019). Personality in a hierarchical model of psychopathology. *Clinical Psychological Science*, *7*, 77–92. https://doi.org/10.1177/2167702618797105

Williams, D. R. (2018). Bayesian inference for gaussian graphical models: Structure learning, explanation, and prediction. *PsyArXiv*. https://doi.org/10.31234/osf.io/x8dpr

Williams, D. R., & Rast, P. (2018). Back to the basics: Rethinking partial correlation network methodology. *PsyArXiv*. https://doi.org/10.31219/osf.io/fndru

Woo, S. E., Chernyshenko, O. S., Longley, A., Zhang, Z.-X., Chiu, C.-Y., & Stark, S. E. (2014). Openness to experience: Its lower level structure, measurement, and cross-cultural equivalence. *Journal of Personality Assessment*, *96*, 29–45. https://doi.org/10.1080/00223891.2013.806328

Wood, D., Gardner, M. H., & Harms, P. D. (2015). How functionalist and process approaches to behavior can explain trait covariation. *Psychological Review*, *122*, 84–111. https://doi.org/10.1037/a0038423

Wright, A. G., & Simms, L. J. (2014). On the structure of personality disorder traits: Conjoint analyses of the CAT-PD, PID-5, and NEO PI-3 trait models. *Personality Disorders: Theory, Research, and Treatment*, *5*, 43–54. https://doi.org/10.1037/per0000037

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*, 1100–1122. https://doi.org/10.1177/1745691617693393

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*, 17. https://doi.org/10.2202/1544-6115.1128

Zimmermann, J., Woods, W. C., Ritter, S., Happel, M., Masuhr, O., Jaeger, U., ... Wright,

A. G. C. (in press). Integrating structure and dynamics in personality assessment: First steps toward the development and validation of a Personality Dynamics Diary. *Psychological Assessment*. https://doi.org/10.31234/osf.io/5zcth

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432–442. https://doi.org/10.1037/0033-2909.99.3.432

Supplementary Information

A Psychometric Network Perspective on the Measurement and Assessment of Personality Traits

## Method

**Design and Data Generation.** A small-scale simulation was performed to evaluate the performance of exploratory graph analysis with ordianl data and whether network measures are linked to traditional psychometric measures. In general, the simulated structure was setup to determine the hierarchy of a single personality domain. Thus, the number of dimensions refers to facets of a single domain (e.g., Extraversion) rather than multiple different domains (e.g., the FFM). The simulation was designed with 2 randomly varying conditions and 3 manipulated conditions. Factor loadings and correlations between dimensions were allowed to vary randomly between values of .30 to .70 (drawn from a uniform distribution). This range of factor loadings is commonly seen in personality inventories (e.g., NEO PI-R; McCrae & Costa, 1987). Similarly, correlations between dimensions (i.e., facets) typically tend to have moderate to large effect sizes (e.g., BFAS; DeYoung, Quilty, & Peterson, 2007). Sample size (500, 1000, and 2500), number of dimensions (i.e., facets; 2, 4, and 6), and number of variables per dimension (6, 8, 10) were manipulated on the basis of commonly used sample sizes in scale development and personality inventories (e.g., NEO PI-R; 8 variables and 6 facets and BFAS; 10 variables and 2 aspects). This resulted in a $3 \times 3 \times 3$ simulation design (27 simulated conditions in total). One hundred datasets per condition were simulated, resulting in 2700 simulated datasets.

All items were generated as continuous items and categorized into five categories (representing a 5-point Likert scale response format). Two additional steps were taken to provide more realistic dimensional structures and data. Cross-loadings were implemented to provide dimension structures that are similar to those found in real data (see Table 1; Bollmann, Heene, KÃ¼chenhoff, & BÃ¼hner, 2015). These were generated following the procedure described in GarcÃa-GarzÃ³n, Abad, & Garrido (in press). The first two items of each dimension were set as markers (i.e., all of their cross-loadings were zero), with the rest

randomly drawn from a normal distribution, *N(0,0.05)*. The skewness was also manipulated for each item, following Garrido, Abad, & Ponsoda (2013). Skewness was randomly drawn with equal probability from a range of -2 to 2 in increments of 0.50. Preliminary checks of several personality datasets previously obtained by the authors were consulted to verify appropriate skewness of the items. Item skewness typically ranged from -1 to 1 for each factor of the FFM. Thus, we included a broader range of skewness to allow for more severe items (i.e., mostly positively or negatively endorsed), which would be expected in pre-selected item pools. Data generation followed the procedures of (Golino et al., 2018, pp. 16–17), including the use of the same data generation code, which was adapted for ordinal data.

## Dimension Identification Methods

**Graphical LASSO.** The graphical LASSO (GLASSO; Friedman, Hastie, & Tibshirani, 2008) is the most commonly applied network construction method in the psychometric network literature. Networks estimated using the GLASSO are a Gaussian Graphical Model (GGM; Lauritzen, 1996), where edges represent partial correlations between variables after conditioning on all other variables in the network. The least absolute shrinkage and selection operator (LASSO; Tibshirani, 1996) is used to control for spurious relationships and shrink coefficients to zero, generating a sparse network and preventing over-fitting of the data. Current applications in psychometric networks use the extended Bayesian information criterion (EBIC; Foygel & Drton, 2010) to select the best fitting GLASSO model (GeLASSO; Epskamp, 2016; Epskamp & Fried, 2018). The EBIC uses a hyperparameter ($\gamma$), which adjusts the penalization and shrinkage of coefficients to zero (the larger $\gamma$, the greater the penalization). Most commonly, $\gamma$ is set to 0.5 (Foygel & Drton, 2010), although greater sensitivity (true positive proportion) can be gained from lower $\gamma$ values (e.g., $\gamma = 0$) at the cost of specificity (true negative proportion; Williams & Rast, 2018). In this simulation, the default $\gamma = 0.5$ was used unless there were disconnected nodes. On these occasions, $\gamma$ was set to 0.25 to maintain sparsity and ensure all nodes were

connected. The GeLASSO in exploratory graph analysis is applied using the *qgraph* package (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012) in R.

**Triangulated Maximally Filtered Graph.** The triangulated maximally filtered graph (TMFG; Massara, Di Matteo, & Aste, 2016) is another network construction method that has been used in the psychometric network literature (e.g., Christensen, Kenett, Aste, Silvia, & Kwapil, 2018; Golino et al., 2018). The TMFG algorithm constructs the network using a structural constraint on the number of zero-order correlations that can be included in the network (i.e., $3n - 6$; where $n$ is the number of variables). Construction begins by identifying the four variables with largest sum of correlations to all other variables and connects them to each other. Then, variables are iteratively added to the network based on the largest sum of three correlations to nodes already in the network. The result is a fully connected network of 3- and 4-node cliques (i.e., sets of connected nodes). This 3- and 4-node clique structure can be directly associated with the inverse covariance matrix (i.e., precision matrix), resulting in a GGM (Barfuss, Massara, Di Matteo, & Aste, 2016). In this simulation, edges were left as zero-order correlations. The TMFG in exploratory graph analysis is applied using the *NetworkToolbox* package (Christensen, 2018) in R.

**Parallel Analysis.** As a comparison, we used two parallel analysis (PA) methods: principal axis factoring (PApaf) and principal component analysis (PApca). These two methods were chosen because they have been extensively evaluated in the literature (e.g., Garrido et al., 2013) and have shown comparable performance with EGA and EGAtmfg in a previous simulation study (Golino et al., 2018). In short, PA generates a larger number of random datatsets, with an equivalent number of cases as the original dataset, by resampling (with replacement) from the original dataset (Horn, 1965). The number of factors (PAF) or components (PCA) whose eigenvalues in the original dataset are greater than the mean of the resampled datasets' are suggested as the dimensional solution.

To derive the dimensions that items belonged to, factor analysis was applied with the dimension solution proposed by PApaf and PApca. The dimension with the largest factor

loading for each item was designated to be the dimension the item belonged to. This item allocation was used to compute the normalized mutual information for the PApaf and PApca methods. Both the parallel analyses and the factor analyses were applied using the *psych* package (Revelle, 2017) in R.

All dimension identification methods (DIM) used polychoric correlations computed by the `cor_auto` function in the *qgraph* package in R. From here, EGA will refer to exploratory graph analysis with the GeLASSO method and EGAtmfg will refer to exploratory graph analysis with the TMFG method.

## Community Detection

**Walktrap algorithm.** Exploratory graph analysis uses the walktrap community detection algorithm (Pons & Latapy, 2006) to determine the number of communities (or dimensions) in the network (Golino & Epskamp, 2017). The walktrap algorithm uses a process known as "random walks" (a stochasic number of steps (or edges) from a certain node), which tend to get "trapped" in densely connected parts of the network. The number of steps can be specified by the user; however, exploratory graph analysis uses the default of 4, which has been shown to be optimal for the number of variables typically used in psychological constructs (i.e., variables $< 100$; Pons & Latapy, 2006).

In the random walk process, the likelihood of a step to another node is determined by the structural similarity between the nodes and the communities, which define a distance. These distances are used in an agglomerative hierarchical clustering algorithm approach, which is then subjected to merging (i.e., merging two clusters that minimize the mean of the squared distances). During the merging process, the adjacent clusters' distances are updated to reflect the new distances between the clusters. Throughout this process, a metric to assess the quality of the partitions is used to help capture community structures at different scales. The walktrap algorithm in exploratory graph analysis is applied using the *igraph* package (Csardi & Nepusz, 2006) in R.

## Centrality Measures

**Community centrality.** The community closeness and community eigenvector centrality measures were used as dimension-level measures to be compared with general factor loadings (i.e., each dimension's loading on a single factor) and scale-measure correlations (i.e., sum scores of the items in each dimension correlated with sum score of all items). General loadings and scale-measure correlations were used as proxies for how central a facet (i.e., dimension) is to an overall factor.

The community closeness centrality is based on the node-wise closeness centrality measure. Node-wise closeness centrality is the reciprocal of each node's average shortest path length ($ASPL_i$; the mean number of edges from the reference node to all other nodes). The $ASPL_i$ is defined as,

$$ASPL_i = \frac{\sum_{i \neq j} d(v_i, v_j)}{n-1}, \tag{1}$$

where $d(v_i, v_j)$ indicates the shortest distance (number of edges) between node $i$ and node $j$ and $n$ indicates the number of nodes in the network. Similarly, the community closeness centrality is the reciprocal of the community's average shortest path length. Instead of a single node, each node's $ASPL_i$ in the community is obtained, and then the mean is computed and its reciprocal taken (eq. 2). Thus, the community closeness centrality (CCC) can be defined as,

$$CCC = \frac{1}{\sum_{i \in \mathbf{c}}^{n} \frac{ASPLi}{n}}, \tag{2}$$

where $\mathbf{c}$ indicates a community and $n$ is the number of nodes in the community. One dataset has demonstrated that the CCC is related to the scale-measure correlation when

using EGAtmfg, but more extensive evidence is needed (Christensen, 2018).

The community eigenvector centrality is adapted from Giscard and Wilson's (2018) cycle centrality (Giscard & Wilson, 2017). The cycle centrality is designed to measure the eigenvector centrality—a measure of a node's direct and indirect connections—for any subset of nodes in the network. Their work extends the eigenvector centrality beyond the individual node to a group of nodes (for mathematical proofs, see Giscard & Wilson, 2018). Because the cycle centrality can be applied to any subset of nodes, it provides a viable solution for computing the eigenvector centrality for communities. Therefore, the community eigenvector centrality is simply the cycle centrality for each community defined in the network.

**Node-wise centrality.** The strength and hybrid centrality were used as item-level measures to be compared to item factor loadings and item-scale (i.e., item-dimension) correlations. The strength centrality is the sum of the connections to a given node. This centrality has been shown to be the most stable of the centrality measures that are commonly used in the literature (e.g., Epskamp, Borsboom, & Fried, 2018) and is considered to be most like factor loadings and item-scale correlations (Hallquist, Wright, & Molenaar, 2019; MÃµttus & Allerhand, 2017). The hybrid centrality permits a single, continuous measure of how central a node's position is in the network (Christensen et al., 2018; Pozzi, Di Matteo, & Aste, 2013). More specifically, it's a composite of the rank-order of several other centrality measures (i.e., betweenness, closeness, strength, and eigenvector):

$$HC = \frac{BC^w + BC^u + LC^w + BC^u + k^u + NS^w + EC^w + EC^u - 8}{8 \times (N-1)}, \tag{3}$$

where $w$ signifies the weighted measure (i.e., correlation strength is considered) and $u$ signifies the unweighted measure (i.e., all edges are set to be equal). Each centrality value is the rank-order value *not* the actual centrality value. $BC$ represents betweenness centrality, $LC$ represents closeness centrality, $NS$ represents node strength, $k$ represents degree (i.e.,

number of connections to a node), and $EC$ represents eigenvector centrality. Note that degree ($k$) and node strength ($NS$) are only included once because the weighted degree *is* node strength.

## Data Analysis

R (R Core Team, 2019) was used for all analyses. The figures were generated using the *ggplot2* package (Wickham, 2016) and the ggpubr package (Kassambara, 2018). Both EGA techniques were applied using the *EGA* package (Golino, 2019). All centrality measures were computed using the *NetworkToolbox* package (Christensen, 2018) and CFA models (including factor loadings) were estimated using the weighted least squares with mean and variance-adjusted standard errors (WLSMV) estimator and computed using the *lavaan* package (Rosseel, 2012). CFA models were fit with respect to the dimensions identified by the exploratory graph analysis techniques rather than the simulated structure. This was done to facilitate direct comparisons between the centrality measures and their traditional counterparts. For each simulated dataset, exploratory graph analysis was applied and the CFA model was fit. If there were disconnected nodes or the CFA model did not converge, then the dataset was discarded, and a new dataset was simulated. In total, 382 datasets were discarded and re-simulated.

**Accuracy and bias of DIM.** Accuracy was defined by three different measures: percentage correct (PC), mean absolute error (MAE), and normalized mutual information (NMI; Table SI1). Mean bias error (MBE) was also computed to determine the extent to which a DIM over- or under-factored.

Table SI1
*Accuracy and Bias Measures*

| Measure | Equation | Description |
|---|---|---|
| Percent Correct (PC) | $\frac{\Sigma C}{N_S} \times 100 \; for \; C\{^{1 \; if \; \hat{\theta} = \theta}_{0 \; if \; \hat{\theta} \neq \theta}\}$ | Average number of times the predicted number of dimensions match the number of dimensions in the population |
| Mean Absolute Error (MAE) | $\frac{\Sigma \mid \hat{\theta} - \theta \mid}{N_s}$ | Consistent deviation away from the dimensions in the population |
| Normalized Mutual Informaton (NMI) | $\frac{-2 \; \Sigma_{i=1}^{C_A} \Sigma_{j=1}^{C_B} N_{ij} \log(\frac{N_{ij} N}{N_{i.} N_{.j}})}{\Sigma_{i=1}^{C_A} N_{i.} \log(\frac{N_{i.}}{N}) + \Sigma_{j=1}^{C_B} N_{.j} \log(\frac{N_{.j}}{N})}$ | Overlap of the item allocations between the population and estimatied dimensions (see a more detailed description below) |
| Mean Bias Error (MBE) | $\frac{\Sigma(\hat{\theta} - \theta)}{N_s}$ | Tendency to over- or under-estimate the number of dimensions in the population |

*Note.* $\hat{\theta}$ *= estimated number of dimensions,* $\theta$ *= population number of dimensions,* $N_s$ *= number of simulated samples,* $C_A$ *= population communities,* $C_B$ *= estimated communities,* $N$ *= confusion matrix (where rows are the population communities and columns are the estimated communities),* $N_{i.}$ *= sum over row* $i$ *of matrix* $N_{ij}$*, and* $N_{.j}$ *= sum over column* $j$ *of matrix* $N_{ij}$*.*

NMI is defined by the information that is shared between the simulated dimensions and dimensions identified by the DIM (Danon, Diaz-Guilera, Duch, & Arenas, 2005). NMI provides a more nuanced analysis of accuracy by quantifying how the items are sorted into dimensions compared to the simulated dimensions. For example, a DIM may provide the same number of dimensions as the simulated dimensions, but some items may be misplaced and appear in a different dimension than is defined by the population structure. Thus, NMI quantifies the item's dimensional correspondence between the simulated and estimated structure.

**Analysis of variance.** We conducted analyses of variance (ANOVAs) to identify how the different conditions affected the accuracy of the dimensionality methods. Each

accuracy measure (except MBE) was used as a dependent variable and the conditions (main effects and interactions) were the independent variables. MBE was not included because its positive and negative values could compensate one another (obscuring any potential effects). The partial eta squared ($\eta_p^2$) was used to assess the magnitude of the effects, with $\eta_p^2$ values of .01, .06, and .14 representing small, moderate, and large effect sizes (Cohen, 1992).

**Associations between centrality and traditional measures.** Spearman's rho rank correlation coefficients were used to assess the associations between all centrality and traditional measures. The community centrality measures—closeness and eigenvector—were computed when there were at least 3 or more dimensions detected by the exploratory graph analysis techniques. This was because the centralness of 2 dimensions is difficult to discriminate (i.e., they are equally central). These measures were correlated with general loadings and scale-measure correlations. The node-wise centrality measures—hybrid and strength—were computed for the network rather than individual connections within their community. These measures were correlated with factor loadings and item-scale correlations.

## Small-Scale Simulation Results

Table SI2
*Dimension Accuracy Results*

| | Sample Size | | | Variables per Dimension | | | Number of Dimensions | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | 500 | 1000 | 2500 | 6 | 8 | 10 | 2 | 4 | 6 | Total |
| | | | | | Percentage Correct | | | | | |
| EGA | 65.7 | **77.2** | 88.2 | **69.0** | **78.3** | 83.8 | **98.0** | 83.2 | **49.9** | **77.0** |
| EGAtmfg | **67.1** | 77.1 | 83.7 | 65.0 | 75.0 | **87.9** | 95.9 | **85.4** | 46.6 | 76.0 |
| PApaf | 32.9 | 56.4 | **88.4** | 66.7 | 59.6 | 51.4 | 89.8 | 55.9 | 32.0 | 59.3 |
| PApca | 64.0 | 68.3 | 76.9 | 59.0 | 71.1 | 79.2 | 97.5 | 77.3 | 34.4 | 69.8 |
| | | | | | Mean Bias Error | | | | | |
| EGA | **-0.08** | **-0.14** | -0.14 | **-0.28** | **-0.06** | **-0.02** | 0.03 | **0.01** | **-0.40** | **-0.12** |
| EGAtmfg | -0.32 | -0.26 | -0.18 | -0.47 | -0.21 | -0.08 | 0.04 | -0.10 | -0.70 | -0.25 |
| PApaf | 1.76 | 0.77 | **-0.01** | 0.40 | 0.85 | 1.27 | 0.13 | 0.84 | 1.54 | 0.84 |
| PApca | -0.19 | -0.30 | -0.24 | -0.44 | -0.21 | -0.08 | **0.00** | -0.15 | -0.58 | -0.24 |
| | | | | | Mean Absolute Error | | | | | |
| EGA | 0.48 | **0.28** | 0.15 | **0.43** | **0.28** | 0.21 | 0.03 | 0.19 | **0.70** | **0.31** |
| EGAtmfg | 0.44 | 0.30 | 0.20 | 0.49 | 0.30 | **0.14** | 0.04 | **0.15** | 0.74 | **0.31** |
| PApaf | 1.79 | 0.82 | **0.12** | 0.55 | 0.89 | 1.29 | 0.13 | 0.87 | 1.72 | 0.91 |
| PApca | **0.43** | 0.34 | 0.24 | 0.47 | 0.31 | 0.23 | **0.02** | 0.23 | 0.75 | 0.34 |
| | | | | | Normalized Mutual Information | | | | | |
| EGA | **0.89** | **0.96** | **0.99** | **0.93** | **0.95** | **0.96** | **0.97** | **0.95** | **0.91** | **0.94** |
| EGAtmfg | 0.84 | 0.92 | 0.96 | 0.88 | 0.91 | 0.92 | 0.94 | 0.91 | 0.87 | 0.90 |
| PApaf | 0.85 | 0.95 | **0.99** | **0.93** | 0.93 | 0.93 | **0.97** | 0.93 | 0.89 | 0.93 |
| PApca | 0.88 | 0.94 | 0.97 | 0.90 | 0.93 | 0.95 | **0.97** | 0.94 | 0.88 | 0.93 |

*Note. Average performance for each condition and across conditions (Total). EGA = exploratory graph analysis with GLASSO, EGAtmfg = exploratory graph analysis with TMFG, PApaf = parallel analysis with principal axis factoring, PApca = parallel analysis with principal component analysis. The best values for each condition are bolded.*

**DIM accuracy and bias.** The marginal table (Table SI2) of the dimension accuracy and bias results demonstrate that EGA was the best all-around method followed by EGAtmfg. Both methods methods performed considerably well and were generally better than the PA methods. Of the two PA methods, PApca outperformed PApaf on all metrics. Notably, PApaf was the most accurate and least biased when the sample size was large. Across the methods, there was a modest tendency to under-factor, except for PApaf, which tended to over-factor. For the exploratory graph analysis methods, EGA performed slightly better with a greater percentage correct (PC = 77.0%), lower bias (MBE = -0.12), and

greater normalized mutual information (NMI = 0.94) than EGAtmfg (PC = 76.0%, MBE = -0.25, and NMI = 0.91). Both techniques had an equivalent mean absolute error (MAE = 0.31). Notably, the differences between the exploratory graph analysis techniques were minimal.

Table SI3

*ANOVA Partial Eta Squared ($\eta_p^2$) Effect Size by Condition and Accuracy Measure*

| | | Sample Size | Variables per Dimension | Number of Dimensions |
|---|---|---|---|---|
| EGA | PC | .01 | .00 | **.07** |
| | MAE | .00 | .00 | .05 |
| | NMI | .05 | .00 | .01 |
| EGAtmfg | PC | .00 | .01 | **.08** |
| | MAE | .00 | .01 | **.09** |
| | NMI | .04 | .00 | .01 |
| PApaf | PC | **.10** | .00 | **.11** |
| | MAE | **.11** | .01 | **.10** |
| | NMI | **.10** | .00 | .02 |
| PApca | PC | .00 | .00 | **.13** |
| | MAE | .00 | .00 | **.12** |
| | NMI | .01 | .00 | .02 |

*Note. EGA = exploratory graph analysis with GLASSO, EGAtmfg = exploratory graph analysis with TMFG, PApaf = parallel analysis with PAF, and PApca = parallel analysis with PCA. PC = percent correct, MAE = mean absolute error, and NMI = normalized mutual information. Bolded values represent moderate effect sizes ($\eta_p^2 = .06$).*

The ANOVAs revealed under which conditions affected the different DIM the most (Table SI3). In general, there were no large effect sizes ($\eta_p^2 = .14$). Similarly, there were no interactions that larger than a small effect size ($\eta_p^2 = .01$), so they were not included in the table. All DIM had a moderate effect in percent correct due to the number of dimensions. Similarly, all DIM except EGA had a moderate affect for MAE. This suggests that the number of dimensions (when correlations between dimensions and factor loadings are varied randomly) has the greatest effect on DIM. PApaf also had moderate effects on all accuracy measures due to sample size (in line with the substantial increases in PC across sample size in Table SI2).

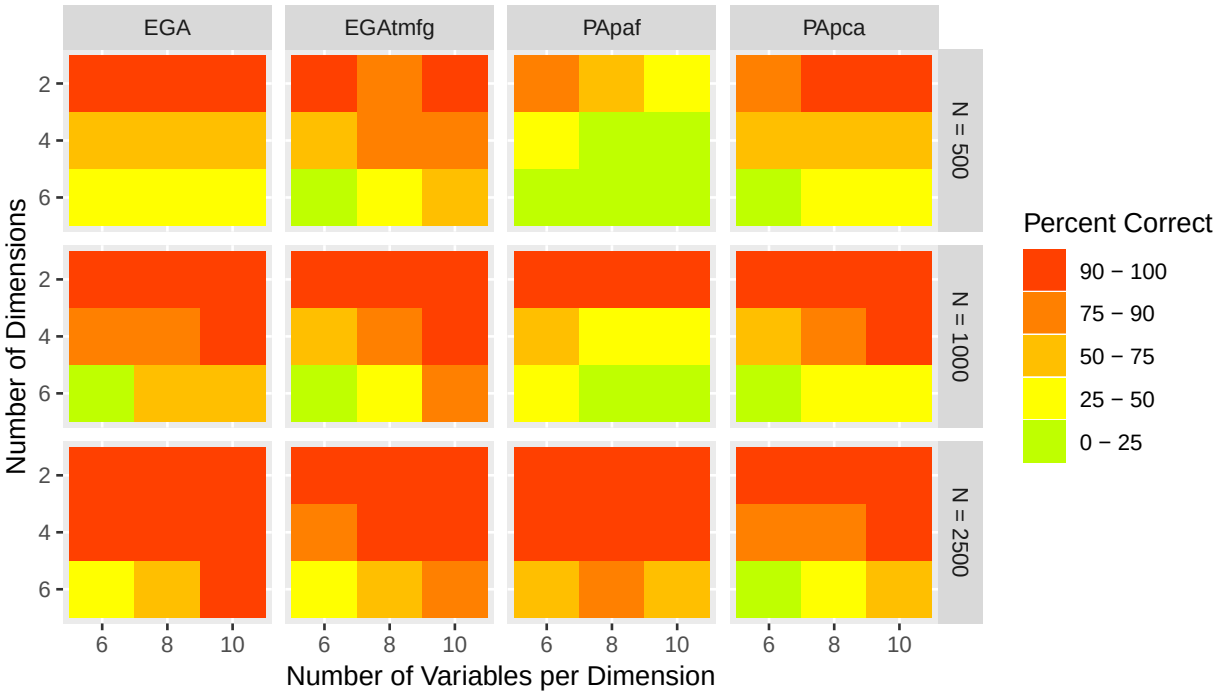


*Figure SI1*. Percent correct plot broken down by each simulation condition set. EGA = exploratory graph analysis with GLASSO, EGAtmfg = exploratory graph analysis with TMFG, PApaf = parallel analysis with PAF, PApca = parallel analysis with PCA. N = sample size.

When broken down by specific condition sets, the results become slightly more nuanced (Figure SI1). In general, the same pattern holds: EGA is more accurate than the rest of the DIM; however, there are a few condition sets where other DIM are more accurate than EGA. For example, when there was a smaller sample size ($N = 500$) or a larger number of variables per dimension (10), EGAtmfg had the best accuracy. Conversely, PApaf performed the best with larger sample sizes ($N = 2500$) and the worst with smaller sample sizes ($N = 500$). PApca was perhaps the most consistent in terms of accuracy across all sample sizes, with accuracy gradually increasing as sample size increased. For most methods, detecting two dimensions or ten variables per dimension was the most accurate, while six dimensions and six variables per dimension was the least accurate.

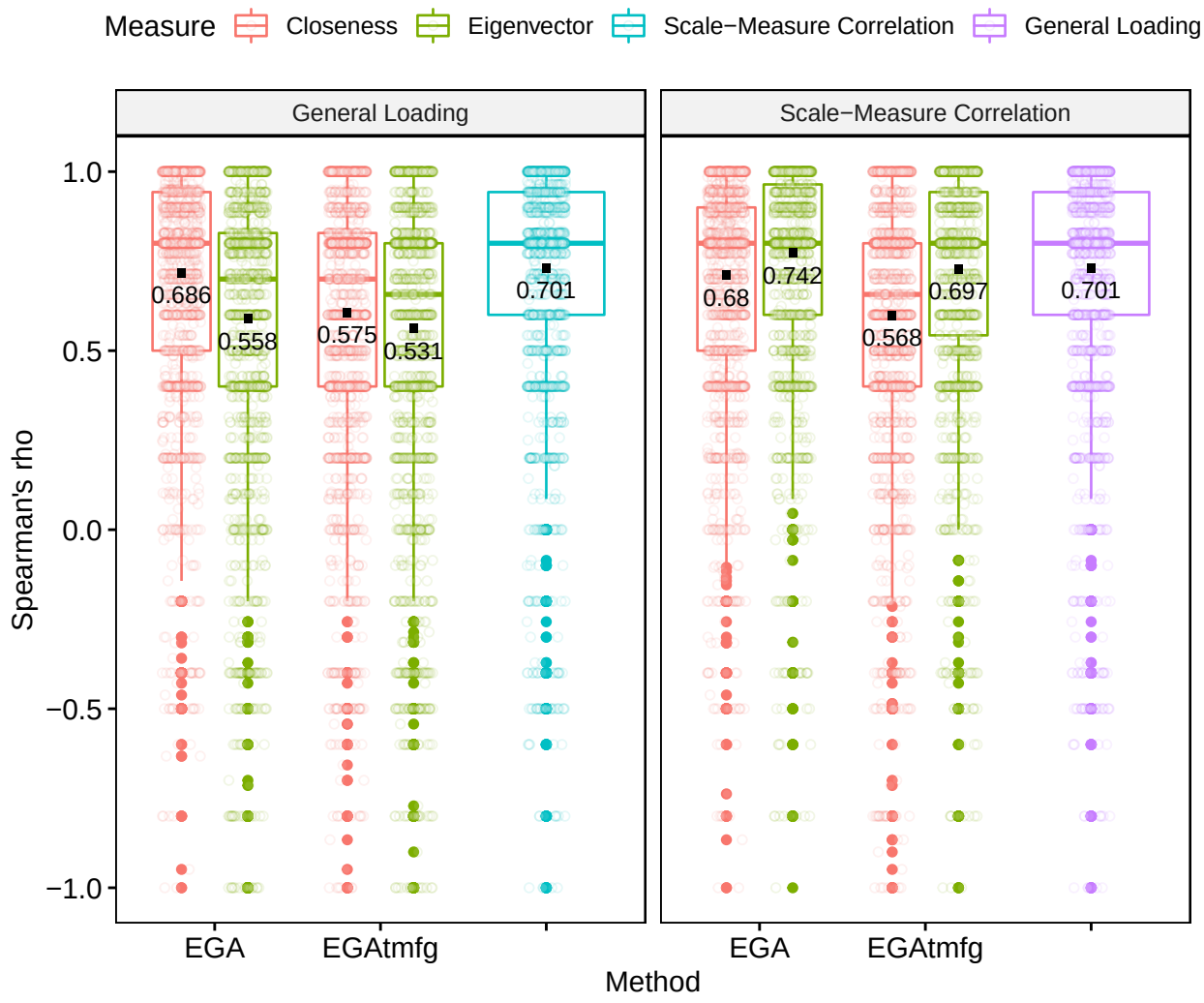


*Figure SI2*. Boxplots of the Spearman's rho between the community measures and traditional measures. The bar across the boxplot represents the median with the upper and lower limits of the boxes representing the upper and lower quartiles. The black dot represents the mean Spearman's rho with the value in the text below. Each transparent dot represents a single simulated sample.

**Dimension measures.** Scale-measure correlations and general loadings had a large Spearman's rho correlation ($\bar{r}$ = .70). Notably, the community centrality measures had comparable effect sizes to the traditional measures (Figure SI2). The CCC had comparable effect sizes for general loadings ($\bar{r}_{EGA}$ = .69 and $\bar{r}_{EGAtmfg}$ = .58) and scale-measure ($\bar{r}_{EGA}$ = .68 and $\bar{r}_{EGAtmfg}$ = .57) correlations between the exploratory graph analysis techniques. The effect sizes for CEC revealed a different pattern: scale-measure correlations ($\bar{r}_{EGA}$ = .74 and

$\bar{r}_{EGAtmfg} = .70$) had larger effect sizes for both exploratory graph analysis techniques than general loadings ($\bar{r}_{EGA} = .56$ and $\bar{r}_{EGAtmfg} = .53$). In general, EGA had larger effect sizes than EGAtmfg for all measures. These results suggest that CCC more closely corresponds general loadings, while CEC more closely corresponds scale-measure correlations. This differentiation suggests that they are measuring similar yet distinct concepts in terms of scale-to-measure weights.

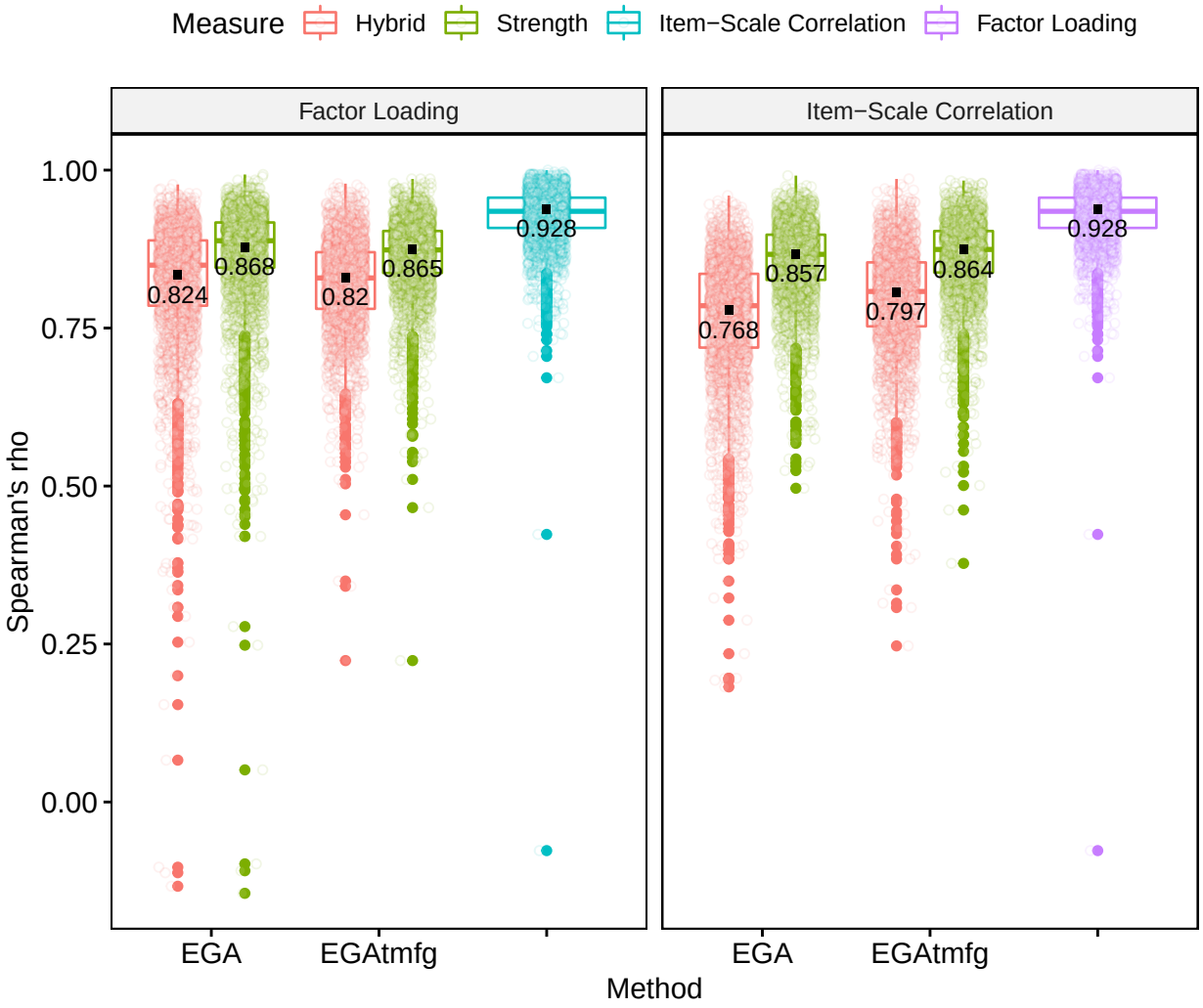


*Figure SI3*. Boxplots of the Spearman's rho between the node-wise centrality and traditional measures. The bar across the boxplot represents the median with the upper and lower limits of the boxes representing the upper and lower quartiles. The black dot represents the mean Spearman's rho with the value in the text below. Each transparent colored dot represents a single simulated sample.

**Item measures.** As expected, the results demonstrate that factor loadings and item-scale correlations were largely redundant ($\bar{r}$ = .93). For the node centrality measures, both centralities were strongly related to factor loadings and item-scale correlations for both exploratory graph analysis techniques, displaying nearly identical effect sizes (Figure SI3). Node strength ($\bar{r}$'s ranging from .85 to .87) had larger average effect sizes than the hybrid centrality ($\bar{r}$'s ranging from .77 to .82) for both exploratory graph analysis techniques. In contrast to the community centrality measures, the node-wise measures had larger effect sizes for EGAtmfg than EGA.

It's important to note, as Hallquist et al. (2019) demonstrate, that centrality measures consider relations that traditional measures do not—connections within and between dimensions are computed in centrality measures, while CFA factor loadings and item-scale correlations are solely within dimension measures. Despite this difference, the centrality measures (particularly node strength) appear to be relatively equivalent to CFA factor loadings and item-scale correlations.

References

Barfuss, W., Massara, G. P., Di Matteo, T., & Aste, T. (2016). Parsimonious modeling with information filtering networks. *Physical Review E*, *94*, 062306. https://doi.org/10.1103/PhysRevE.94.062306

Bollmann, S., Heene, M., KÃ¼chenhoff, H., & BÃ¼hner, M. (2015). *What can the real world do for simulation studies? A comparison of exploratory methods.* Retrieved from Department of Statistics, University of Munich: Retrieved from https://epub.ub.uni-muenchen.de/24518/

Christensen, A. P. (2018). NetworkToolbox: Methods and measures for brain, cognitive, and psychometric network analysis in R. *The R Journal*, *10*, 422–439. https://doi.org/10.32614/RJ-2018-065

Christensen, A. P., Kenett, Y. N., Aste, T., Silvia, P. J., & Kwapil, T. R. (2018). Network structure of the Wisconsin Schizotypy Scales–Short Forms: Examining psychometric network filtering approaches. *Behavior Research Methods*, *50*, 2531–2550. https://doi.org/10.3758/s13428-018-1032-9

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*, 1–9. Retrieved from https://pdfs.semanticscholar.org/1d27/44b83519657f5f2610698a8ddd177ced4f5c.pdf

Danon, L., Diaz-Guilera, A., Duch, J., & Arenas, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, *2005*, P09008. https://doi.org/10.1088/1742-5468/2005/09/P09008

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*, 880–896. https://doi.org/10.1037/0022-3514.93.5.880

Epskamp, S. (2016). Regularized Gaussian psychological networks: Brief report on the

performance of extended BIC model selection. *arXiv*. Retrieved from https://arxiv.org/abs/1606.05771

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*, 617–634. https://doi.org/10.1037/met0000167

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

Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*, 1–18. https://doi.org/10.18637/jss.v048.i04

Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, & A. Culotta (Eds.), *Advances in neural information processing systems* (pp. 604–612). Retrieved from http://papers.nips.cc/paper/4087-extended-bayesian-information-criteria-for-gaussian-graphical-models

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*, 432–441. https://doi.org/10.1093/biostatistics/kxm045

GarcÃa-GarzÃ³n, E., Abad, F. J., & Garrido, L. E. (in press). Improving bi-factor exploratory modeling: Empirical target rotation based on loading differences. *Methodology*.

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods*, *18*, 454–474. https://doi.org/10.1037/a0030005

Giscard, P.-L., & Wilson, R. C. (2017). Cycle-centrality in economic and biological networks. In C. Cherifi, M. Cherifi, M. Karsai, & M. Musolesi (Eds.), *International conference on complex networks and their applications* (pp. 14–28). New York, NY: Springer.

https://doi.org/10.1007/978-3-319-72150-7_2

Giscard, P.-L., & Wilson, R. C. (2018). A centrality measure for cycles and subgraphs II. *Applied Network Science*, *3*, 1–15. https://doi.org/10.1007/s41109-018-0064-5

Golino, H. F. (2019). *EGA: Exploratory Graph Analysis – estimating the number of dimensions in psychological data*. Retrieved from https://github.com/hfgolino/EGA

Golino, H. F., & Epskamp, S. (2017). Exploratory Graph Analysis: A new approach for estimating the number of dimensions in psychological research. *PloS ONE*, *12*, e0174035. https://doi.org/10.1371/journal.pone.0174035

Golino, H., Shi, D., Garrido, L. E., Christensen, A. P., Nieto, M. D., Sadana, R., & Thiyagarajan, J. A. (2018). Investigating the performance of Exploratory Graph Analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *PsyArXiv*. https://doi.org/10.31234/osf.io/gzcre

Hallquist, M., Wright, A. C. G., & Molenaar, P. C. M. (2019). Problems with centrality measures in psychopathology symptom networks: Why network psychometrics cannot escape psychometric theory. *PsyArXiv*. https://doi.org/10.31234/osf.io/pg4mf

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185. https://doi.org/10.1007/BF02289447

Kassambara, A. (2018). *ggpubr: 'ggplot2' based publication ready plots*. Retrieved from https://CRAN.R-project.org/package=ggpubr

Lauritzen, S. L. (1996). *Graphical models*. Oxford, UK: Clarendon Press.

Massara, G. P., Di Matteo, T., & Aste, T. (2016). Network filtering for big data: Triangulated Maximally Filtered Graph. *Journal of Complex Networks*, *5*, 161–178. https://doi.org/10.1093/comnet/cnw015

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*, 81–90. https://doi.org/10.1037/0022-3514.52.1.81

MÃµttus, R., & Allerhand, M. (2017). Why do traits come together? The underlying trait

and network approaches. In V. Ziegler-Hill & T. K. Shackelford (Eds.), *SAGE handbook of personality and individual differences: The science of personality and individual differences* (pp. 1–22). London, UK: SAGE Publications.

Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, *10*, 191–218. https://doi.org/10.7155/jgaa.00185

Pozzi, F., Di Matteo, T., & Aste, T. (2013). Spread of risk across financial markets: Better to invest in the peripheries. *Scientific Reports*, *3*, 1665. https://doi.org/10.1038/srep01665

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Revelle, W. (2017). *psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/ 10.18637/jss.v48.i02

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288. Retrieved from jstor.org/stable/2346178

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.

Williams, D. R., & Rast, P. (2018). Back to the basics: Rethinking partial correlation network methodology. *PsyArXiv*. https://doi.org/10.31219/osf.io/fndru