ON THE EQUIVALENCY OF FACTOR AND NETWORK LOADINGS

Alexander P. Christensen
Department of Psychology
University of North Carolina at Greensboro
Greensboro, NC, 27402
alexpaulchristensen@gmail.com

Hudson Golino
Department of Psychology
University of Virginia
Charlottesville, VA, 22903
hfg9s@virginia.edu

October 9, 2020

ABSTRACT

Recent research has demonstrated that the network measure *node strength* or sum of a node's connections is roughly equivalent to confirmatory factor analysis (CFA) loadings. A key finding of this research is that node strength represents a combination of different latent causes. In the present research, we sought to circumvent this issue by formulating a network equivalent of factor loadings, which we call *network loadings*. In two simulations, we evaluated whether these network loadings could effectively (1) separate the effects of multiple latent causes and (2) estimate the simulated factor loading matrix of factor models. Our findings suggest that the network loadings can effectively do both. In addition, we leveraged the second simulation to derive effect size guidelines for network loadings. In a third simulation, we evaluated the similarities and differences between factor and network loadings when the data were generated from random, factor, and network models. We found sufficient differences between the loadings, which allowed us to develop an algorithm to predict the data generating model called the *Loadings Comparison Test* (LCT). The LCT had high sensitivity and specificity when predicting the data generating model. In sum, our results suggest that network loadings can provide similar information to factor loadings when the data are generated from a factor model and therefore can be used in a similar way (e.g., item selection, measurement invariance, factor scores).

Keywords psychometric networks · node strength · factor loadings

1   Introduction

In most areas of psychology, reflective latent variable models are the standard conceptualization of measurement (Borsboom, Mellenbergh, & van Heerden, 2003). In a reflective model, observable variables that measure a psychological attribute are thought to co-occur because of an underlying common cause—that is, an unobserved (latent) attribute causes the covariation between the observed variables (often referred to as the common cause model; Marsman et al., 2018; Schmittmann et al., 2013). This perspective is most often provided by factor analysis in which reflective latent variables (or factors) are pointing to the observed variables (Edwards & Bagozzi, 2000).

In the last decade, mutualism models have emerged as an alternative explanation for the formation of psychological attributes proposing that observable variables co-occur because they reciprocally and dynamically reinforce one another, forming a causally connected system (Borsboom, 2008; van der Maas et al., 2006). This perspective is most often provided by network models, which are depicted by nodes (circles) representing variables (e.g., psychopathological symptoms) and edges (lines) representing the partial correlation between two nodes conditioned on all other nodes (Epskamp & Fried, 2018).

Despite their differing representations and hypotheses about the cause of the co-occurrence between observable variables, more recent research has demonstrated that latent variable and network models can be mathematically equated (Golino & Epskamp, 2017; Hallquist, Wright, & Molenaar, 2019; Marsman et al., 2018; van Bork et al., 2019). In fact, Guttman (1953) explicated this equivalence nearly 60 years ago when proposing a generalized dimensional approach called *image structural analysis*, which regresses all variables on each other and does not rely on the introduction of hypothetical variables (i.e., latent variables). Although network models were not yet formalized in psychology, Guttman's approach is essentially the basis of contemporary node-wise regression network models where each node in the network is regressed on all other nodes (Haslbeck & Waldorp, 2020).

Guttman noted in his exposition that the Spearman-Thurstone common factor theory is a special case of his image theory in which error terms are uncorrelated. Therefore, if the data generating mechanism is a factor model, the variance-covariance matrix of the *anti-images* (i.e., correlations between the residuals) tends toward a diagonal matrix as the number of items(n) increases (and the off-diagonal elements are equal to zero in the limit as n approaches infinity), irrespective of the number of factors.

A consequence, as Guttman (1953, p. 296) points out, is that "if not all non-diagonal elements of the [variance-covariance matrix of the anti-images] are close to zero, one is led to reject the hypothesis that a determinate common-factor space of rank less than n holds for the universe." In other words, Guttman's image structural analysis could be the first step in checking the plausibility of the existence of underlying latent factors in multivariate data. The connection between image theory and factor analysis thus goes in the same direction of the assumption of conditional independence in factor models where variables are uncorrelated given a hypothetical latent variable (Sobel, 1997).

A separate implication of this equivalency is that different data generating mechanisms, such as those suggested by latent variable (factor) and network models, can create similar sets of statistical relations (Marsman et al., 2018; van Bork et al., 2019). As van Bork et al. (2019) point out, this similarity of statistical relations extends into the first (means) and second (variance-covariance matrix) moments, which means that any covariance matrix can be represented as a factor and network model. Guttman (1953) shows that the covariance that is uniquely shared between each pair of variables supports a more general form of common factor theory. Thus, although the models are representing different covariance partitions of the variance-covariance matrix (zero-order vs. partial correlations), they are still deriving information from the same set of statistical relations and therefore should provide the similar information (e.g., Waldorp & Marsman, 2020).

The goal of our paper is to further evaluate the equivalence between factor and network models by showing that network models can estimate an equivalent metric to factor loadings. Deriving such a measure would psychometrically link factor and network models, providing evidence that these models can yield similar information despite offering distinct data generating hypotheses for what this information substantively means. Factor loadings are fundamental to modern psychometrics, providing information about how well a variable is measuring a certain construct and whether a variable is influenced by multiple latent causes (DeVellis, 2017). This substantive interpretation changes when placed into the context of network models, which propose that variables co-occur not because of latent causes but because they are causally coupled. This data generation hypothesis provides the substantive interpretation that factors in a network model "emerge" from their constituent causal connections rather than cause them (Cramer, 2012). This suggests that the interpretation of a network equivalent factor loading would be each node's unique contribution to the emergence of a coherent *dimension* (or collection of related variables whose relations are not necessarily due to a common cause; Christensen et al., 2020).[1]

Recent research has demonstrated that one particular network measure, *node strength* or the sum of a node's connections, is statistically redundant with confirmatory factor analysis (CFA) factor loadings (Hallquist et al., 2019). The aim of our study is to build on this work by developing a novel modification of node strength, which we refer to as *network loadings*. The purpose of these network loadings is to circumvent the limitation of *latent confounding* or multiple causes on a variable that was pointed out by Hallquist et al. (2019). Moreover, we evaluate whether network loadings provide different information from factor loadings when data are generated from a model other than a factor model (e.g., random and network models). We begin by reviewing Hallquist et al.'s (2019) three simulation studies to motivate the development of a network equivalent to factor loadings.

---

[1]Throughout the paper, we use factor analytic terminology to explain and report our findings even when referring to network models. This is done to maintain clarity and conciseness of our discussion without being bogged down by constantly stating that network models hold different data generating hypotheses. It's important to note, however, that a factor *could* be interpreted as a collection of variables that are co-occurring because of reciprocal causal relations, and this collection of co-occurring variables in a network *could* be interpreted as occurring because of a common cause (Christensen et al., 2020; Golino, Shi, et al., 2020). Ultimately, it is the researcher's choice of psychometric model and the interpretation of that model that dictates how they theoretically view these relations occurring (Borsboom, 2006; Fried, 2020; Marsman et al., 2018).

2 Review of Hallquist, Wright, and Molenaar (2019)

Across Hallquist and colleagues' (2019) three simulation studies, their goal was to investigate how *centrality measures* or the quantification of a node's relative position in a network were related to and affected by latent influences on variables. Centrality measures are the most frequently used statistic to quantify networks in psychology and to date have been pointed to as a key difference between what latent variable and network models measure (Bringmann et al., 2019; Cramer, Waldrop, van der Maas, & Borsboom, 2010). The three centrality measures they used in their study were *betweenness*, *closeness*, and *node strength* (often referred to as strength). Conceptually, high betweenness nodes are ones that are most often used on the shortest path (i.e., shortest number of edges) from one node to another, high closeness nodes are ones that have the shortest average number of edges to all other nodes in the network, and high node strength nodes are ones that have the largest (absolute) sum of their weighted edges.

As Hallquist et al. (2019) point out, these centrality measures are secondary statistics because they depend on the structure of the network, which must first be estimated. One common approach has been to estimate a Gaussian Graphical Model (GGM; Lauritzen, 1996) where the edges between nodes are partial correlations conditioned on all other nodes in the network. This is typically achieved by applying the *graphical least absolute shrinkage and selection operator* (GLASSO; Friedman, Hastie, & Tibshirani, 2008, 2014) and using the extended Bayesian information criteria (EBIC; Chen & Chen, 2008) to select the best fitting model (Epskamp et al., 2018b). The network model estimated by the GLASSO is often *sparse* meaning that many nodes are not connected to one another (Epskamp et al., 2017a). Centrality measures are then used to estimate the relative positions of nodes in the network. Consequently, the validation of centrality findings are specific to the network estimation method. For all of their simulation studies, Hallquist et al. (2019) used the GLASSO model with EBIC model selection to estimate their networks.

In their first simulation study, they examined whether there was any correspondence between betweenness, closeness, and node strength centrality measures and CFA factor loadings in unidimensional and multidimensional factor models. Their results demonstrated that betweenness and closeness centrality were highly correlated with the CFA loadings of the one factor model ($r = .74$ and $r = .94$, respectively) but had much lower correlations with these loadings when there was more than one factor (r's between .31 and .55). In the one factor condition, they found that the variability in correlations for both measures were due to sampling variability, while in the multiple factor conditions the low correlations were due to sampling variability and the measures dependence on the connectedness of the entire network and therefore were more affected by cross-factor associations than the dominant factor loading. In contrast, node strength was significantly correlated with the factor loadings across all conditions (r's between .97 and .98). Because of the lack of correspondence of betweenness and closeness centrality with factor loadings, we focus solely on node strength for the rest of their simulations.

In their second simulation study, they examined the effects of common versus specific sources of covariation or the extent to which two indicators on different factors were related through a shared separate factor (these will be referred to as the target indicators). These effects were examined in one of the target indicators and a comparator indicator (i.e., an indicator on the same factor as the respective target indicator). In contrast to the first simulation, there was only one condition with two factors and all but one variable (the target indicator) in their respective factors had a factor loading of .80. The two target indicators that shared a separate factor had their correlation vary between $r = 0$ and $r = .64$. This allowed them to examine the extent to which node strength was associated with the variance from the target indicators' respective factors (specific variance) to their shared factor (shared variance).

A general finding of their second simulation was that the edge weight (i.e., partial correlation) between the target indicators had a nearly perfect relationship with the extent to which there was a specific association between them ($r = .997$). As for the node strength estimates, there was a moderate main effect of *specific-to-shared variance balance* (or the difference of the specific and shared variance) and a large main effect of indicator type (target and comparator). This suggests that there was a large increase in a node's strength due to the shared factor. The comparator indicator's node strength had a small main effect from the specific-to-shared variance balance, suggesting minimal impact from the shared factor. Importantly, they noted that there was a nonlinear (cubic) relationship for node strength's specific-to-shared variance balance.

In their final simulation study, they examined the effects of multiple latent causes. This study was setup with a two-factor model with eight indicators per factor and the target indicator that loaded onto both factors (i.e., 17 indicators in total). The target indicator had factor loadings on both factors ranging between .20 and .80 in increments of .05. All other loadings were fixed at .80. Like their second simulation, they also examined a comparator indicator. The results of this study revealed that the target indicator's node strength was an equally weighted combination of factor 1 and factor 2 loadings (both r's = .94). The comparator indicator's node strength was weakly associated with the variation of the target's factor loadings on factor 1 and factor 2.

In sum, their simulations demonstrated that node strength was roughly redundant with CFA factor loadings and affected by different causal sources. These takeaways are each important for their own reasons. The first finding suggests that there is a strong connection between node strength and factor loadings, which means that node strength could potentially be used as an equivalent psychometric tool as factor loadings. The second finding suggests that the relationship between node strength and factor loadings should be tempered in a way that reflects the unique causes in the system. This suggests that the unique causal components of a network must be identified before network measures can be meaningfully interpreted (or at least interpreted similarly to factor loadings; see Christensen et al., 2020 for how this can be achieved; Hallquist et al., 2019).

3 Present Research

We had three goals for the present research. For our first goal, we wanted to develop a modification of node strength that could account for its limitations uncovered by Hallquist and colleagues' second and third simulations (i.e., network loadings). Specifically, we wanted to explicitly test how node strength relates to multiple latent causes when split by those causes, which we achieved by replicating Hallquist and colleagues' second simulation. We expected that doing so would circumvent the confounding effects of different latent causes that affect node strength's interpretation (i.e., where the "strength" of each node is coming from).

For our second goal, we used a Monte Carlo simulation approach to examine the accuracy of the network loadings in estimating population factor loadings across a variety of different data conditions. The use of population loadings contrasts with Hallquist and colleagues' first simulation where node strength was correlated with CFA loadings. A direct comparison with the population factor loadings is a better benchmark for whether network models can accurately identify this information. We also computed EFA and CFA loadings to evaluate the extent to which network loadings were more similar to EFA or CFA loadings. For Simulations 1 and 2, we generated continuous and polytomous data to evaluate the effects of discrete categories on network loadings.

For our third goal, we wanted to examine whether factor and network loadings differed when the data generating model was a random, factor, or network model. For this goal, we first compared the descriptive statistics of the loading proportions that were greater than or equal to small, moderate, and large effect sizes as well as dominant and non-dominant (i.e., cross-) loadings in data generated from random, factor, and network models. Based on these proportions, we then derived heuristics for determining whether the data were generated from a random, factor, or network model. Finally, we employed a third simulation to evaluate the effectiveness of these heuristics in an algorithm designed to predict the true data generating model.

4 Simulation 1: Effects of a Shared Latent Cause

Our first simulation set out to determine whether the network loadings could circumvent the issue of multiple latent causes that was found for the traditional node strength measure. To test this goal, we directly replicated Hallquist and colleagues' second simulation using our network loadings.

5 Methods

5.1 Data Generation

We generated data from multivariate normal factor models following the same approach as Golino, Shi, et al. (2020). First, the reproduced population correlation matrix was computed:

$$\mathbf{R_R} = \mathbf{\Lambda \Phi \Lambda}',$$

where $\mathbf{R_R}$ is the reproduced population correlation matrix, lambda ($\mathbf{\Lambda}$) is the k (variables) $\times$ r (factors) factor loading matrix, and $\mathbf{\Phi}$ is the $r \times r$ correlation matrix. The population correlation matrix, $\mathbf{R_P}$, was then obtained by putting the unities on the diagonal of $\mathbf{R_R}$. Next, Cholesky decomposition was performed on the correlation matrix such that:

$$\mathbf{R_P} = \mathbf{U}'\mathbf{U}.$$

If the population correlation matrix was not positive definite (i.e., at least one eigenvalue ≤ 0) or any single item's communality was greater than 0.90, then $\mathbf{\Lambda}$ was re-generated and the same procedure was followed until these criteria are met. Finally, the sample data matrix of continuous variables was computed:

$$\mathbf{X} = \mathbf{ZU},$$

where $\mathbf{Z}$ is a matrix of random multivariate normal data with rows equal to the sample size and columns equal to the number of variables. To generate polytomous data, each continuous variable was categorized into five categories, resembling a 5-point Likert scale, with a random skew ranging from -2 to 2 on a 0.5 interval from a random uniform distribution following the approach of Garrido, Abad, and Ponsoda (2011, 2013).

5.2 Psychometric Network Model

The GGM was used as the psychometric network model. The GGM is a network model where nodes represent variables and edges represent the partial correlation between two nodes given all other nodes in the network. The GLASSO has been the most commonly applied GGM network estimation method in the network psychometrics literature (Epskamp & Fried, 2018). The *least absolute shrinkage and selection operator* (LASSO; Tibshirani, 1996) of the GLASSO is a statistical regularization technique that reduces parameter estimates with some estimates becoming exactly zero (for mathematical notation, see Epskamp & Fried, 2018). The aim of this technique is to achieve a sparse model—non-relevant edges are removed from the model, leaving only a subset of statistically relevant (but not necessarily significant) edges (Epskamp et al., 2017a).

The popular approach in the network psychometrics literature is to compute models across several values of λ (usually 100) and to select the model that minimizes the EBIC (Epskamp & Fried, 2018). The EBIC model selection uses a hyperparameter (γ) to control how much it prefers simpler models (i.e., models with fewer edges; Foygel & Drton, 2010). Larger γ values lead to sparser models, while smaller γ values lead to denser models. When $\gamma = 0$, the EBIC is equal to the Bayesian information criterion. In the psychometric network literature, this approach has been termed EBICglasso and is applied via the qgraph package (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012) in R (R Core Team, 2020). For continuous data, Pearson's correlations were computed; for polytomous data, polychoric correlations were computed.

Following the exploratory graph analysis (EGA) network estimation approach (Golino, Shi, et al., 2020), the minimum λ value is set to .1, which is slightly larger than the default of .01. This larger value is selected to reduce the possibility of false positive edges in the network. Second, the γ value is set to .50, which is the default; however, it is iteratively decreased by .25 until reaching zero based on whether there is a single node that is disconnected from the network (i.e., it has no connections to any nodes). If γ reaches zero, then the network is used regardless of whether any nodes are disconnected.

The EGA approach also applies a community detection algorithm (Golino, Shi, et al., 2020) to determine the number of communities (or factors) in the network. Because we are only interested in the performance of the network loadings, we assigned items to their population factors to avoid confounding our results with the process of estimating the factors. The *EGAnet* package (Golino & Christensen, 2020) in R was used to compute networks following the EGA approach.

5.3 Network Loadings

An important finding of Hallquist and colleagues' (2019) simulations was that node strength represented a combination of dominant and cross-factor loadings. To circumvent this issue, a node's strength can be split between the nodes in each factor. This can be mathematically written as:

$$S_i = \sum_{j=1}^{n} |w_{ij}|,$$

$$L_{if} = \sum_{j \in f}^{F} |w_{ij}|,$$

where $|w_{ij}|$ is the absolute weight (e.g., partial correlation) between node i and j, S_i is the sum of the edge weights connected to node i across all nodes (n; i.e., node strength for node i), L_{if} is the sum of edge weights in factor f that are connected to node i (i.e., node i's loading for factor f), and F is the number of factors (in the network). This measure can be standardized using the following formula:

$$z_{L_{if}} = \frac{L_{if}}{\sqrt{\sum L_{.f}}},$$

where the denominator is equal to the square root of the sum of all the weights for nodes in factor f. Notably, the standardized loadings are absolute weights with the signs being added after the loadings have been computed (following the same procedure as factor loadings; Comrey & Lee, 2013). In contrast to factor loadings, the network loadings are computed after the number of factors have been extracted from the network's structure. Variables are deterministically assigned to specific factors via a community detection algorithm rather than the traditional factor analytic standard of their largest loading in the loading matrix. This means that some nodes may not have any connections to nodes in other factors in the network, leading some variables to have zeros for some factors in the network loading matrix.

These standardized network loadings summarize the information in the edge weights and so they depend on the type of association represented by the edge weight (e.g., partial correlations, zero-order correlations). Thus, the meaning of these network loadings will change based on the type of correlation used. Specifically, partial correlations represent the unique measurement of a factor whereas zero-order correlations represent the unique and shared contribution of a variable's measurement of a factor. The network loadings were computed using the `net.loads` function in the *EGAnet* package.

5.4 Design

Following Hallquist and colleagues' (2019) second simulation design, we generated two orthogonal factors (no cross-loadings) with ten variables per factor. All except one variable on each factor (i.e., the target indicators) had factor loadings of .80. The target indicators loaded on to their respective orthogonal factors but also a third shared factor. The loadings for these indicators were manipulated such that their variance explained by the shared factor ranged from 0% to 64% in 1% increments: $r^2_{factor} + r^2_{shared} = .64$, where r^2_{factor} is their respective factors and r^2_{shared} is their shared factor. This gave us a total of 65 population models that were replicated 500 times with a sample of 400 for each model replication.

6 Results

The goal of our first simulation was to determine how the network loadings were affected by multiple latent causes. More specifically, we wanted to test whether the network loadings could dissociate the variance of separate and shared latent causes on two target variables. We assigned the target variables to a third factor across all conditions and computed the network loadings. By doing this, we could specifically determine the magnitude of the loading on each variable's respective factor (i.e., separate latent causes) as well as the shared factor (i.e., shared latent cause) by obtaining the corresponding values. We ignored any cross-loadings that were for the factor that the target variables were not assigned to. We computed Spearman's rho correlations to quantify the association between the magnitude of each network loading and their corresponding variance explained values for both the respective and shared factors.

As depicted in Figure 1, there was a strong relationship between the network loadings and the variance of both respective (separate) and shared latent causes, suggesting that the network loadings could distinctly dissociate the latent causes of their respective and shared factors.
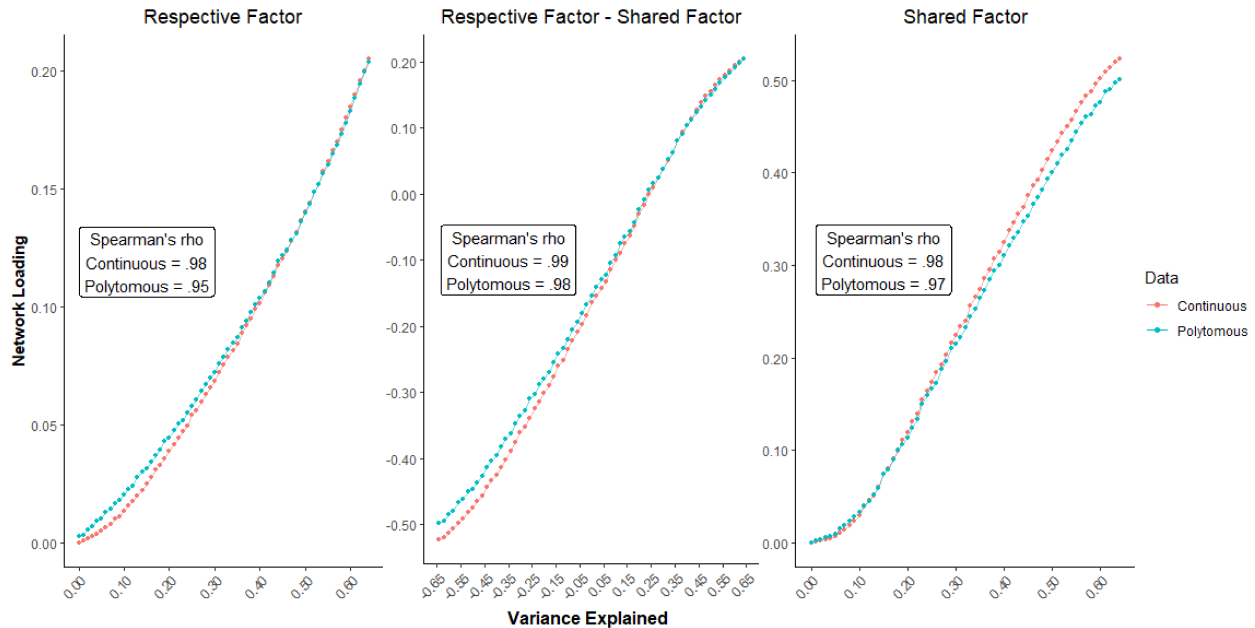
Figure 1: The left figure depicts the relationship between the means of the network loadings (circles) for each population variance condition for the target variables loading onto their respective (separate) factors. The middle figure depicts the relationship between the means of the network loadings (circles) for each population variance condition for the target variables loading onto their respective (separate) factors substracted from the shared factor. Negative values represent greater network loadings and variance explained for the shared factor, while positive values represent greater network loadings and variance explained for their respective factors. The right figure depicts the relationship between the means of the network loadings (circles) for each population variance condition for the target variables loading onto the shared factor.

For the continuous data, the correlations were both $r = .98$ for the respective and shared factors; for the polytomous data, the correlations were $r = .95$ and $r = .97$ for the respective and shared factors, respectively. Following Hallquist and colleagues (2019), we subtracted the shared factor's explained variance from the respective factor variance resulting in what they referred to as specific-to-shared variance balance. Similarly, we subtracted the target indicators' shared network loadings from their respective factor network loadings resulting in what we call *differential network loadings*. We found that there was a strong linear relationship between these differential network loadings and the specific-to-shared variance (middle of Figure 1). Indeed, when regressing the specific-to-shared variance on the differential network loadings, the relationship was strongly linear for both the continuous (Adj. $R^2 = 0.960$) and polytomous data (Adj. $R^2 = 0.938$) with little additional variance being contributed by the quadratic (Adj. $R^2 = 0.965$ and Adj. $R^2 = 0.944$, respectively) and cubic (Adj. $R^2 = 0.970$ and Adj. $R^2 = 0.946$, respectively) terms. These findings contrast with the cubic relationship that Hallquist et al. (2019) found with the traditional node strength measure and suggest that the network loadings circumvented the issue of a confounding latent factor. In short, the network loadings could effectively dissociate the variance from different latent causes.

7   Simulation 2: Correspondence between Factor and Network Loadings

In Simulation 1, we verified that network loadings can account for different latent causes on a variable, which circumvents the issue identified in the traditional node strength measure. The goal of our second simulation was to extend the effectiveness of network loadings to a more general context by building on Hallquist and colleagues' (2019) first simulation where they compared node strength to CFA factor loadings.

Importantly, our simulation expanded on their study in two substantive ways. First, our study compared the accuracy of network, CFA, and EFA loadings in the estimation of population factor loadings as well as to each other. This contrasts with their study where node strength was correlated with CFA loadings only. Comparing network loadings with CFA and EFA loadings allows for a better comparison of what network loadings are more "like." On the one hand, CFA loadings typically offer a simple structure where indicators only load on their factor. On the other hand, EFA loadings offer a "saturated structure" in the sense that each variable has a dominant and cross-loading on every factor. Network

loadings offer a complex structure in-between these two structures where some variables have cross-loadings, while variables may not be associated with certain factors resulting in some variables with a loading of zero on those factors.

Second, our simulation generated data from more complex conditions and data structures. Specifically, we generated data with dominant loadings randomly varying between .40 and .70 and manipulated a broader range of correlations between factors ranging from orthogonal (.00) to large (.70). In addition, cross-loadings were randomly drawn from a normal distribution with its variance increasing as the correlations between factors increased. In contrast, Hallquist and colleagues generated data with a simple structure (without cross-loadings). Finally, we manipulated sample size (very small to large) as well as generated continuous and polytomous data.

As a secondary aim, we identified the correspondence between the population and network loadings. Because the network loadings are derived from partial correlations, the typical guidelines corresponding to small, moderate, and large factor loadings (.40, .55, and .70, respectively; Comrey & Lee, 2013) are not immediately transferable to network loadings. To get a better grasp on the correspondence between the magnitude of the population and network loadings, we leveraged the large number of samples generated in this simulation to provide effect size guidelines for network loadings using the cut-offs from traditional factor analysis.

8 Methods

8.1 Data Generation

We generated data from multivariate normal factor models following the same procedure as Simulation 1.

8.2 Loadings

8.2.1 CFA

For the CFA model, we used the *lavaan* package's (Rosseel, 2012) `cfa` function to estimate factor loadings. The CFA models were specified with the known population factor structure of the data—that is, the population factors with the items placed in their known factors. For the continuous data, we used the maximum likelihood estimator; for the polytomous data, we used the weighted least square mean and variance adjusted estimator.

8.2.2 EFA

For the EFA model, we used the *psych* package's (Revelle, 2017) `fa` function to estimate the factors in the data. Because the number of factors is known, we specified the population number of factors as the number of factors to compute in the EFA. The factor model was estimated using the maximum likelihood for continuous data and weighted least squares for polytomous data. For both types of data, we used the geomin oblique rotation from the *GPArotation* package (Bernaards & Jennrich, 2005).

8.2.3 Network

The network loadings were estimated using the same GLASSO and network loading procedure described in Simulation 1. Similar to the CFA model, nodes were assigned to the known population structure of the data to avoid confounding the evaluation of the network loadings with the performance of community detection algorithms (e.g., Christensen & Golino, 2020).

8.2.4 Factor orientation

To orient the variables so that they were in the same structural pattern as the population factors, we computed the absolute sum of each factor's loadings in the estimated methods corresponding to the variables in the population factors. The factor with the largest absolute sum was then assigned as the dominant factor for those variables. This orientation was done for each loadings method. We excluded samples from our analyses in the instance that the same estimated factor was selected for more than one population factor (1.8% of samples in total across both continuous and polytomous data).

8.3 Design

The population models were simulated from a multidimensional multivariate normal distribution with two, three, and four factors, and four, eight, and twelve variables per factor. The dominant factor loadings on each factor randomly

varied between .40 and .70. The correlations between factors were orthogonal (.00), small (.30), moderate (.50), and large (.70). As the correlations between factors increased, the cross-loadings increased. The cross-loadings were generated from a random normal distribution with a mean of zero and standard deviation of .050, .075, .100, and .125 for orthogonal, small, moderate, and large correlations between factors, respectively. This generation of cross-loadings led to both positive and negative values.

With each step increase in variance, the magnitude of the cross-loadings also increased. This allowed some cross-loadings to achieve nearly equivalent magnitudes as the dominant loadings in the large correlations between factors condition (.70). This adjustment provides a practical implementation of our first simulation where some variables are the result of multiple latent causes, which is increasingly likely with larger correlations between factors. Finally, sample sizes of 250 (very small), 500 (small), 1000 (moderate), and 5000 (large) were generated.

This simulation design allowed for a mixed factorial design: $3 \times 3 \times 4 \times 4 \times 2$ (number of factors $\times$ variables per factor $\times$ correlations between factors $\times$ sample size $\times$ number of responses) for a total of 288 simulated condition combinations and 72,000 simulated datasets.

8.4   Statistical Analyses

To compare the performance of the CFA, EFA, and network loadings, we used Spearman's rank-order correlation between each method's loadings and the known population loadings. Rank-order rather than Pearson's correlations were chosen because they have a larger penalty for loadings that differ in their order from the population loadings. Moreover, we computed ANOVAs to evaluate the main and interaction effects of conditions on the correlations between the population loadings. Following Cohen (1992), we used partial eta-squared (η_p^2) effect sizes of small (.01), moderate (.09), and large (.25). Finally, we computed Spearman's rank-order correlation between all loadings to evaluate their relations to each other and to determine whether the network loadings were more related to EFA or CFA loadings.

9   Results

Across all conditions, the EFA loadings were the most accurate ($\bar{r} = .90$) followed by the network ($\bar{r} = .88$) and CFA loadings ($\bar{r} = .82$; Figure 2). As a general trend, all loading estimation methods were strongly affected by the number of factors ($\geq .15$; see SI 2). The decrease in accuracy with the number of factors was most pronounced for the CFA loadings ($\eta_p^2 = 0.91$), which were likely affected by the simple structure of the loading matrix.
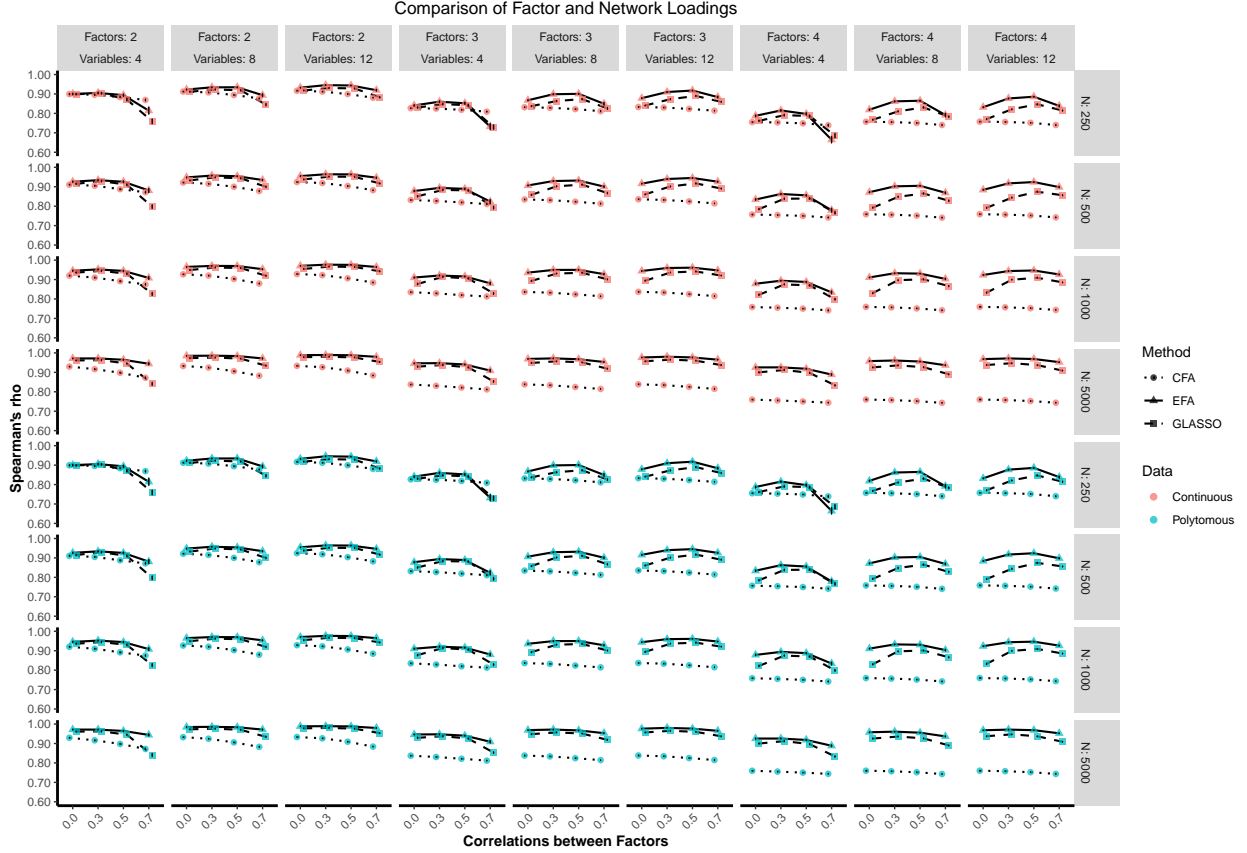
Figure 2: Comparison of factor and network loadings broken down by each condition. Error bars represent the 95 percent confidence intervals are represented for each mean. Network loadings (GLASSO) are represented by the dashed line and square, CFA loadings are represented by the dotted line and circle, and EFA loadings are represented by the solid line and triangle. Continuous data are presented in red and polychoric data are presented in blue. Note that the y-axis begins at .60.

Another general trend was that the network loadings tended to follow a similar pattern to the EFA loadings (Figure 2), which was corroborated by their correlations across data types ($\hat{r}_{continuous} = .92$, $\hat{r}_{polytomous} = .88$, and $\hat{r}_{overall} = .90$). Similarly, both EFA and network loadings had equivalent effect sizes when correlated with CFA loadings for continuous and polytomous data (both $\hat{r}$'s = .83 and .81, respectively) as well as across both data types (both $\hat{r}$'s = .82). A more detailed breakdown of these results can be found in the Supplementary Materials (SI 1).

9.1 Correspondence between Population Factor and Network Loadings

To get a better understanding of how the population factor and network loadings aligned, we visualized their correspondence by rounding the population factor and network loadings to two decimal places and using absolute values to obtain greater power for these values (Figure 3). As shown in Figure 3, the network loading magnitudes for the continuous and polytomous data did not differ by much in their correspondence to the population factor loadings suggesting that these magnitudes were consistent regardless of data type.

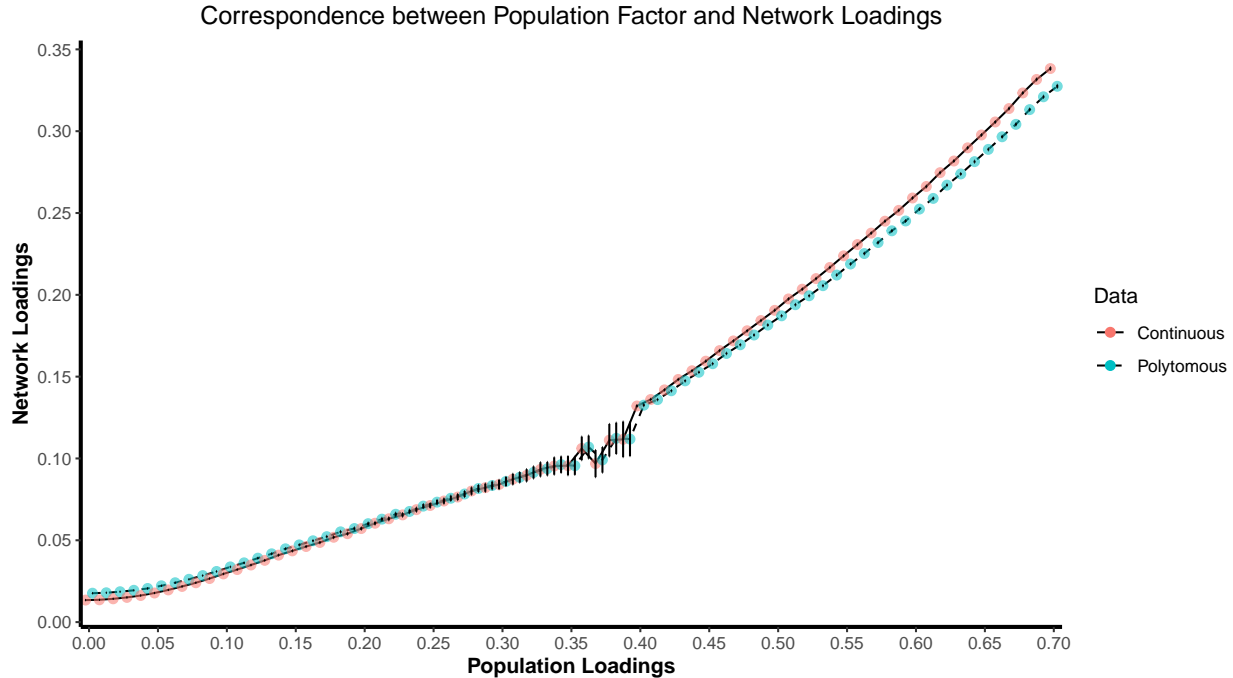Correspondence between Population Factor and Network Loadings

Figure 3: Correspondence between the population factor and network loadings. The circles represent the mean network loading for the continuous (red) and polytomous (blue) data. Error bars represent the 95 percent confidence intervals are represented for each mean.

There were three population loading effect size magnitudes of interest: 0.40, 0.55, and 0.70. For the population loading of 0.40, the average network loading magnitude was 0.13 for both continuous and polytomous data. For the population loading of 0.55, the average network loading magnitude was 0.22 for both continuous and polytomous data; for the population loading of 0.70, the average network loading magnitude was 0.34 and 0.33 for the continuous and polytomous data, respectively. Taking these results in hand, we suggest general effect size guidelines for network loadings to be 0.15 for small, 0.25 for moderate, and 0.35 for large. Interestingly, there were strong linear trends for the network loadings up to and after the population loading magnitudes between 0.35 and 0.40 with the network loading magnitudes being the most variable within that range. This variability does not suggest that the network loadings are inconsistent in this range but rather is an effect of a small number of samples with population factor loadings in this region (2.7% of the total number of population loadings across all samples).

10 Simulation 3: Different Data Generating Models

So far, we've shown that network loadings can circumvent the issue of latent confounding and effectively estimate the population factor loadings across a variety of conditions. Moreover, we've shown that there is a strong linear correspondence between network and factor loadings, which we used to derive effect size guidelines for the network loadings. In short, our findings demonstrate that network loadings are roughly equivalent to factor loadings when the data generating model is a factor model. This leaves open an important question: What if the data are generated from a different model?

One key difference in the statistical basis of factor and network models provide some hints about what should be expected. Factor models attempt to extract the common covariance between variables within a certain number of factors. This means that the magnitude of factor loadings depends on the extent to which there exists enough shared variance across sets of variables. If, for example, data are generated from a random model (random independent variables), then the common covariance in the data should be sparse and diffuse. This is because there is no common cause across variables and therefore little shared variance. When attempting to extract a certain number of factors, there should be relatively few small, moderate, or large factor loadings because the common covariance between any set of variables is scarce, which makes the chances of larger loadings small.[2]

---

[2]It's important to note that we are conceptualizing a random model as a specific model: a random correlation matrix with small correlations centered around zero (i.e., random independent variables whose covariance is due to sampling variability). If data are

Network models do not extract the common covariance between variables but instead map the unique associations between the variables with the number of factors being estimated afterwards. That is, network loadings are the standardized sum of each node's connection to a specified factor meaning that their magnitudes only depend on each node's relative contribution to the overall sum of the weights in the factor. From a conceptual point of view, the network perspective suggests that components of the network should be "causally autonomous" such that they "differ [in their] causes and effects of other components." (Cramer et al., 2012, p. 415). In this sense, a random correlation matrix could be a network model such that each variable *is* causally autonomous. This should lead network loadings to produce at least some small, moderate, and large effect sizes even when the model is given a certain number of factors. On this basis, factor loadings may be more informative than network models for determining whether data were generated from a random model because the network loadings will still provide differential information about the relations between variables and factors.

Although network models may be substantively more aligned with a random model than factor models, most real-world phenomena are not random. Indeed, many real-world networks tend to have a *small-world* structure, which is between a completely regular (lattice) structure where each node is connected to a certain number of neighbors (i.e., neighborhood) and a completely random structure where there is no inherent pattern to each node's connections (Watts & Strogatz, 1998). To be more specific, small-world networks are characterized by nodes having many neighboring connections but also some cross-network connections with even fewer nodes that act as *hubs* or nodes with an above average number of connections. These "small-world" characteristics have been found in psychopathological symptoms where any two symptoms in the network were on average about three symptoms away from each other (Borsboom, Cramer, Schmittmann, Epskamp, & Waldorp, 2011). These characteristics provide some inference into what might be expected when factor or network loadings are estimated for data generated from a small-world network model.

For factor models, the connectivity between neighbors suggests that there is likely to be some common covariance within the clusters (i.e., sets of connected nodes) of the network while the cross-network connectivity suggests that there is likely to be some covariance between these clusters. Moreover, nodes that act as hubs should allow a greater chance of some moderate and large factor loadings relative to a random model. In a general sense, these characteristics seem to resemble a factor model with moderate correlations between factors. A key difference between a factor and small-world network model is that being connected to a neighbor does not necessarily suggest that they are all highly correlated with one another (as they would in a factor model) but rather that they are simply connected. This means that although neighbors may be connected, their connections to other nodes in the network could be stronger than with each other.

Based on this notion, factor models would likely find enough common covariance across all variables to effectively extract one factor but the residual common covariance would likely be more diffuse across any other factors, leading to larger loadings on mostly one single factor with fewer moderate and large loadings on other factors. For network models (and network loadings by extension), much of this common covariance is removed yielding the unique associations between variables. These unique associations would then be partitioned across any factors detected, which would likely lead to fewer large network loadings but an abundance of small to moderate loadings relative to factor models. Moreover, network loadings would likely have a greater proportion of cross-loadings that are of at least a small effect size. These differences between the two loadings may be informative for determining whether data were generated from a factor or network model.

The goal for this simulation was twofold. First, we generated data from random, factor, and network models to qualitatively evaluate whether our theoretical rationale fit the data. Our focus was on examining the descriptive statistics of the proportion of loadings that were greater than or equal to small, moderate, and large effect sizes as well as dominant and cross-loadings greater than or equal to a small effect size for both network and factor loadings.

The rationale for using the proportion of loading effect sizes are to summarize the data structures such that no matter the number of factors or variables there are an equivalent number of parameters used comparison. By using proportions that are equal to or larger than a certain effect size, more continuous cut-offs are used that reduce some of arbitrariness that is inherent in rule-of-thumb effect sizes. Dominant and cross-loading proportions were particularly important because we expected that factor loadings would have larger dominant loadings when data were generated from a factor model relative to the other models. Similarly, we expected more network cross-loadings that were at least a small effect size when data were generated from a network model relative to the other models.

Based on these proportion statistics, we then sought to specify heuristics that could be used to determine whether data were generated from a random, factor, or network model. Using these heuristics, we derived an algorithm called the *Loadings Comparison Test* (LCT), which we then investigated using a Monte Carlo simulation. In this simulation, we generated 15 variables from random, factor, and network models. In total, we generated 6000 samples from each

---

generated from a random correlation matrix where correlations between variables are large, then this assumption does not hold. That is, there will be large factor loadings because there is more shared variance. A factor model will, however, fit poorly because such a matrix would not be low rank. We thank the anonymous reviewer for pointing this out to us.

model and we manipulated conditions of the factor (correlations between factors) and network (rewiring probability and neighborhood size) models. To evaluate the LCT, we computed proportion correct and confusion matrix metrics for whether the algorithm could correctly predict the appropriate data generating model.

11   Methods

11.1   Data Generation

In our initial simulation to obtain descriptive statistics (hereafter referred to as the *descriptive simulation*), 1000 samples were generated from each model and their associated conditions. In the simulation testing our algorithm (hereafter referred to as the *testing simulation*), we generated 6000 samples for each model, which were divided equally among the conditions within each model (6000 for random, 1500 per condition for factor, and 400 per condition for network). All data were generated using continuous values. Details about the descriptive simulation method can be found in the Supplementary Information (SI 2).

11.1.1   Random model

Fifteen variables were generated from a random normal distribution with a mean of zero and standard deviation of five. Next, we computed the variance-covariance matrix between these variables, performed a Cholesky decomposition this matrix, and computed the inverse of the matrix. After, we multiplied the matrix by the original data to force the variables to have correlations of zero. A separate 15×15 correlation matrix was generated from a random normal distribution with a mean of 0 and standard deviation of 0.10 (i.e., zero-order correlations generally ranged from -.20 to .20). The diagonal of this correlation matrix was set to 1 and a Cholesky decomposition was performed. This matrix was then multiplied by the fifteen variables to obtain our final data values. There were no conditions to manipulate for the random model.

11.1.2   Factor model

We generated data from multivariate normal factor models following the same procedure as Simulations 1 and 2. For all factor models, we generated three factors with five variables per factor (15 variables in total). We only manipulated the correlations between factors (.00, .30, .50, and .70). As in our previous simulations, the variance of the cross-loadings were also adjusted with the magnitude of correlations between factors (.050, .075, .100, .125, respectively).

11.1.3   Network model

For all network models, fifteen variables were generated from a GGM small-world network model using the `genGGM` function in the *bootnet* package (Epskamp et al., 2018a) following the approach of Yin and Li (2011) and Epskamp et al. (2017b), and the small-world network algorithm of Watts and Strogatz (1998). This procedure starts with a network structure without weights and then draws weights from a uniform distribution. In order to generate models that are more closely aligned with data observed in psychology, we had weights drawn from a uniform distribution ranging from 0.1 and 1 that were made negative with a 30% probability. In addition, we set the diagonal elements of the matrix to 0.8 (in contrast to the default of 1.5) times the sum of all absolute values in the corresponding row. These adjustments increased the maximum weight (partial correlation) from around 0.16 to 0.32 while also increasing the average non-zero weights from 0.08 ($SD = 0.04$) to 0.16 ($SD = 0.07$) for data generated from a small-world network model with 15 variables, neighborhood of 4, and rewiring probability of 0.50. The *neighborhood* of a small-world network refers to the number of nodes that each node is connected to in the initial lattice structure. The *rewiring probability* refers to the probability that any one node's connection would be randomly connected to another node in the network. We manipulated both neighborhood (3, 4, and 5) and rewiring probability (0, 0.25, 0.50, 0.75, and 1). The neighborhood values were chosen to represent the sparsest possible small-world network with 15 variables, which require number of variables (n) > neighborhood > $ln(n)$ > 1 (Watts & Strogatz, 1998). The network densities corresponding to the neighborhood values of 3, 4, and 5 were 0.43, 0.57, and 0.71, respectively.

11.2   Loadings

For all loadings, absolute values were used. We computed EFA loadings following the same procedure as Simulation 2. The factor model was estimated using the maximum likelihood and oblimin rotation. The network loadings were estimated using the same GLASSO and network loading procedure described in Simulation 1.

11.3 Estimating Factors

A key part of obtaining and qualitatively comparing the factor and network loadings was to estimate the same number of factors for both. To do so, we estimated the number of factors using the default EGA approach, which estimates the GLASSO with EBIC model selection (as described in Simulation 1) and applies the Walktrap algorithm (Pons & Latapy, 2006). The Walktrap algorithm begins by computing a transition matrix where each element represents the probability (based on node strength) of one node traversing to another. Using Ward's agglomerative clustering approach (Ward, 1963), each node starts as its own cluster and merges with adjacent clusters (based on squared distances between each cluster) in a way that minimizes the sum of squared distances between other clusters. A statistic called *modularity* (or the extent to which nodes have more connections within their respective cluster and fewer connections with other clusters; Newman, 2006) is used to determine the optimal partition of clusters (i.e., factors).

The Walktrap algorithm, as implemented in *igraph* (Csardi & Nepusz, 2006), is deterministic meaning that it always returns the same number of factors and allocation of variables in those factors given a network structure. For the network model, the number of factors and item allocations were used to compute the network loadings. The allocation of variables were used to determine dominant and non-dominant (i.e., cross-) loadings. For the factor model, the number of factors returned by the algorithm was then used as the number of factors estimated in the EFA model. The variables of the EFA model were then assigned to the factor which had their largest absolute loading. This allowed us to define loadings as dominant or non-dominant.

11.4 Statistical Analyses

We computed measures based on the confusion matrix of the algorithm accurately predicting the true data generating model. To provide an example, we use the random model as the model under consideration. A true positive (TP) was when the predicted and true generating model matched the model under consideration (e.g., random). A true negative (TN) was when the predicted and true generating model (e.g., factor or network) were not the model under consideration (e.g., random). A false positive (FP) was when the predicted generating model matched the model under consideration (e.g., random) but not the true generation model (e.g., factor or network). A false negative (FN) was when the predicted generating model (e.g., factor or network) did not match the model true generating model and model under consideration (e.g., random).

Using this confusion matrix, we computed sensitivity ($\frac{TP}{TP+FN}$), specificity ($\frac{TN}{TN+FP}$), false discovery rate (FDR; $\frac{FP}{FP+TP}$), accuracy ($\frac{TP+TN}{TP+FP+TN+FN}$), and Matthews correlation coefficient (MCC; $\frac{(TP \times TN)-(FP \times FN)}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}}$). Sensitivity is the proportion of TPs that are correctly identified as TPs, while specificity is the proportion of TNs that are correctly identified as TNs. The FDR is the proportion of FPs that are found relative to the total positives that are predicted by the algorithm. Accuracy is the proportion of correct predictions (TPs and TNs) of the algorithm, representing an overall summary of sensitivity and specificity. Finally, the MCC is considered to the best overall metric for classification evaluation because it is an unbiased measure that uses all aspects of the confusion matrix, representing a special case of the phi coefficient between the predicted and true model (Chicco & Jurman, 2020).

12 Results

We present and discuss the results of the descriptive simulation as well as the heuristics we used for the LCT algorithm in the Supplementary Information (SI 2).

12.1 Testing Simulation

Using the LCT, we performed a testing simulation to evaluate its performance. The algorithm is deterministic meaning that it will provide a single prediction for the data generating model. The effectiveness of the LCT was contextualized using the percent of correctly predicted models within each condition (Table 1) and confusion matrix metrics across all conditions (Table 2).

Table 1: Loadings Comparison Test Percent Correct by Condition

Model	Condition	Rewiring Probability	Percent Correct (%)
Random			95.7
Factor	.00		100
	.30		100
	.50		98.6
	.70		77.9
Network	3	0	88.2
		.25	91.5
		.50	95.8
		.75	90.2
		1	92.2
	4	0	87.8
		.25	90.8
		.50	94.0
		.75	94.8
		1	95.0
	5	0	94.8
		.25	93.2
		.50	93.8
		.75	94.2
		1	95.5

Note. For the factor models, condition is the correlation between factors; for the network models, condition is the neighborhood.

As shown in Table 1, the LCT performed very well across all models and conditions. The lowest accuracy (77.9%) for the LCT was for factor models with large correlations (.70) between factors. This is not surprising given that the factor models were most descriptively like the network models (see SI 2). Statistically, factor models with such large correlations between factors are more likely to appear like network models because they will tend to have cross-loadings that create a covariance structure that is similar to networks. To get a better understanding of how well the LCT performed as an overall predictive model, we examined its confusion matrix (Table 2).

Table 2: Loadings Comparison Test Confusion Matrix

	Model		
	Random	Factor	Network
Sensitivity	0.957	0.941	0.928
Specificity	0.994	0.970	0.949
FDR	0.013	0.059	0.099
Accuracy	0.981	0.961	0.942
MCC	0.958	0.911	0.871

Note. All models had 6000 samples in total. FDR = false discovery rate and MCC = Matthews correlation coefficient.

Across all confusion matrix metrics, the LCT performed very well. The LCT had high sensitivity and specificity for all models (all > .90). These metrics are further reflected in the accuracy where the proportion of correctly predicted (TPs) and correctly not predicted (TNs) models were above .90. When considering FDR, random models were rarely

predicted other than when they were the true model. The value of 0.013 suggests that out of 100 models predicted to be a random model only one was not actually a random model. For factor models, about 6 of 100 models that were predicted to be a factor model were not actually a factor model. For network models, about 1 in 10 models that were predicted to be a network model were actually not. The higher FDR for network models was largely driven by factor models with large correlations between factors (.70) where all incorrect LCT predictions were network models (22.1%). Indeed, when this condition is removed, the network's FDR is cut in half (0.05 or 1 in 20 models that were predicted to be a network model were actually not). Finally, the MCC was large across all models. Because the MCC is equivalent to the phi coefficient, its effect size can be interpreted the same (Powers, 2011). Therefore, the values can be interpreted as correlation coefficients making them substantially large effect sizes across models. In sum, the LCT was highly effective at predicting the data generating model.

13   Discussion

This study sought to derive and evaluate a modified node strength measure (i.e., network loadings) that separated specific contributions of latent causes. Our study builds on the results and recommendations from Hallquist et al.'s (2019) simulation studies. In our first simulation, we demonstrated that when node strength is modified to be split between latent causes it is no longer affected by latent confounding and can sufficiently account for multiple latent causes. In our second simulation, we submitted these network loadings to a more general context where they were evaluated across a variety of data conditions and compared against EFA and CFA loadings. In large part, our results demonstrate that these network loadings can accurately recover the population loadings of a factor model and are more closely related to EFA loadings. Finally, we evaluated whether network and factor loadings remained similar when the data generating model was not a factor model. We found that the loadings differed enough to derive a highly sensitive and specific algorithm for predicting the data generating model.

As a general takeaway, it's important to point out that the factor and network loadings were not a strict one-to-one equivalency but rather provided roughly equivalent information when the generating model was a factor model. This was made most clear in Simulation 2 and 3 where network loadings were roughly equivalent to factor loadings only when the data generating model was a factor model. When the model was a random model, the factor loadings did not have much information (in regard to the magnitude of loading) relative to network loadings. When the data was a small-world network model, the factor loadings were much more alike a factor model; however, the cross-loadings appeared to be restricted relative to the network loadings.

Although factor and network loadings are similar when the data generating model was a factor model, it's the selection and representation of the model that implies a specific causal interpretation (Borsboom, 2006; Bringmann & Eronen, 2018). It's plausible that a factor actually represents an emergent property of a collection of variables while it's equally plausible that a collection of variables in a network represents a common cause (Guyon, Falissard, & Kop, 2017; Kruis & Maris, 2016; van der Maas et al., 2006). For factor models, their *representation* implies that a latent variable causes the relations between variables; for network models, their *representation* implies that the variables co-occur through causal relations and that dimensions emerge from these relations (Bringmann & Eronen, 2018; Christensen et al., 2020; Marsman et al., 2018). Therefore, the equivalence of factor and network loadings (in factor models) is of relatively no consequence *statistically*, but matters *theoretically*.

This difference in substantive interpretation is what largely separates factor models from network models (Bringmann & Eronen, 2018; Christensen et al., 2020; Marsman et al., 2018). Factor loadings refer to how well an indicator measures an underlying common cause; network loadings refer to the coupling of components that lead to the emergence of "factors" in a causal system. In this sense, network loadings represent each node's (unique) contribution to the emergence of a coherent dimension (or collection of related variables) in the network (Christensen et al., 2020). From this perspective, it is the researcher's hypothesis about the data generating mechanisms that should align with the psychometric model they select, which in turn shapes the way they interpret the statistics of their model. This does not mean that researchers cannot use a network model to identify the latent factor structure of their data (e.g., Golino, Shi, et al., 2020), but it does suggest that researchers must be wary of what their model implies (Borsboom, 2006). In general, the LCT provides researchers with a starting point for what model may be most appropriate for their data, allowing for more informed decisions about how their data should be interpreted.

13.1   Implications

The results from our simulations have several implications for the use of network models in measurement and assessment. One implication comes from our third simulation, which demonstrates how researchers could structure the relationships of their data to mirror their data generating hypothesis. This means that for network models variables should be unique and causally autonomous to avoid latent confounding (Hallquist et al., 2019) but also to better align with network theory

(Christensen et al., 2020; Cramer et al., 2012). Thus, the LCT could be used, for example, to determine whether the structure of a researcher's assessment instrument agrees with their data generating hypothesis and provide direction for when it does not.

Another implication extending from our first and second simulation is that network loadings can be used in an equivalent way as factor loadings, providing many of the same measurement opportunities as factor models. A network loading matrix, for example, can be derived and used for item selection in scale development and validation (DeVellis, 2017). Results from our second simulation identified effect size guidelines that correspond with traditional factor loading guidelines: small (0.15), moderate (0.25), and large (0.35) network loadings. These guidelines should be of particular utility for applied researchers and those validating scales from the network perspective (for a review, see Christensen et al., 2020). Our findings also open the door to computing measurement invariance measures such as metric equivalence of network loadings.

In addition, we suggest that network loadings can be used to derive a weighted between-person score for each participant (or the network equivalent of factor scores). This implication in particular requires more detailed attention. Specifically, how should a network score be computed and substantively interpreted? When considering a network of extraversion components, for example, the network itself references the state of the system—that is, the extent to which the network is in an extraverted state (Christensen et al., 2020). From this perspective, extraversion represents a summary statistic of how variables in the network are influenced by one another (Cramer, 2012). Therefore, a network score is more analogous to a formative latent variable (i.e., a weighted composite) than a reflective latent variable (i.e., common covariance). This substantive explanation suggests that a network score should be computed as a weighted composite, which could be derived from each dimension's sum of the product between its respective network loadings and each person's corresponding variable scores (Guttman, 1953, Theorem 3) . Indeed, in a simulation capturing dynamic factor scores, these network scores correlated with factor scores between .90 and .99 in dichotomous and continuous data (Golino et al., 2020).

13.2 Limitations

The computation of the network loadings possesses a few limitations. First, the computation starts with the network estimation method (e.g., GLASSO), which determines the structure of the network (i.e., how the variables are connected to one another). From this structure, a community detection algorithm (e.g., Walktrap) is applied to estimate the number of factors in the network as well as which items belong to those factors. The network loadings for each variable are then computed based on how the items were allocated. In a random model, for example, spurious communities that reflect sampling variability may be identified and therefore network loadings would reflect loadings onto spurious factors.

This means that the computation of the network loadings is affected by both the network estimation method and community detection algorithm. For the GLASSO, this means that for smaller sample sizes (e.g., 250) there may be fewer connections in the network than when the sample size is large (e.g., 1000). Although larger edges are likely to remain, smaller values in the larger sample may not be included in the network of the smaller sample and therefore will not be included in the computation of the network loadings, effectively lowering the reliability of the estimates. Similarly, the community detection algorithm's accuracy for the number of dimensions and placement of nodes will be affected. Specifically, they will be less reliable when the sample size is small (Golino, Shi, et al., 2020). In addition, small loadings or large cross-loadings will also affect the accuracy of the algorithm's placement of variables into factors (Christensen & Golino, 2020), which in turn affects the reliability of the estimates.

These effects all flow downstream into the computation of the network loadings. For small sample sizes, it's likely that the estimates for the network loadings will be smaller than the population loadings due to sparser information in the network. Small sample sizes and less-defined factor structures will also lead divergences away from population loadings of a factor model through less accurate placements of items. Moreover, different placements of items into factors will also affect the magnitude of the loadings. Therefore, the computation of network loadings depends on the accurate estimation of the network's structure and the factors that are obtained from that structure.

Another limitation is the potential for one data generating model to produce a data structure that is more similar to another data generating model. It's possible, for example, to generate data from a factor model that has large correlated residuals, which has a data structure that is more like a network model than a factor model. Similarly, generating a random correlation matrix with large correlations between variables will produce a structure that is more like a dense network model.[3] Thus, the parameters of each of these models could be manipulated in a way to reflect a data structure that corresponds to other data generating models. In this way, the LCT can be used as a tool to better understand the most likely data generating mechanism given the data's structure, but must be tempered with theoretical expectations.

---

[3]Indeed, when generating random correlation matrices ($n = 3,000$) with correlations between 0.60 and 0.80, the LCT suggests that a network model is the data generating model for three-quarters of the samples (76.4%) and a random model for the rest.

Finally, the LCT is based on descriptive heuristics from a limited set of conditions. We think that it's unlikely that these set of heuristics will generalize across all data structures and conditions. We view the LCT as a promising first step towards a tool for predicting the underlying data generating mechanism. Future work should consider machine learning methods that are able to better approximate the underlying data generating function than the human-based heuristics used in the current implementation (Hastie, Tibshirani, & Friedman, 2017). Such methods may also provide a more continuous prediction about the data generating model (e.g., likelihood the model is one model or another) rather than the single prediction made by the current implementation.

13.3   Conclusion

In conclusion, the statistical equivalency between factor and network models is well documented (Golino & Epskamp, 2017; Guttman, 1953; Marsman et al., 2018; van Bork et al., 2019). Our simulation studies provide differential evidence of this equivalency by evaluating factor and network loadings in random, factor, and network models. We argue that the statistical equivalency in factor models is of (relatively) no consequence and it is the representation and theoretical implications of these models that matters. We echo calls from many others that urge researchers to think critically about the psychometric model they use and the data generating mechanisms that they imply (Borsboom, 2006). After all, "the mapping from statistical association structure to a generating causal structure is typically one-to-many" (Marsman et al., 2018, p. 26), which suggests that it is the representation and interpretation of the model that provides meaning to equivalent statistical structures of the data. Finally, we provide an algorithmic test to predict the model generating the data so that researchers can make more informed decisions about which model may be most appropriate and whether the data's structure matches their expectations. Future work should continue to evaluate the extent to which factor and network models are equivalent.

14   References

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, *65*, 676–696. https://doi.org/10.1177/0013164404272507

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., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PLoS ONE*, *6*, e27407. https://doi.org/10.1371/journal.pone.0027407

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

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., . . . Snippe, E. (2019). What do centrality measures measure in psychology networks? *Journal of Abnormal Psychology*, *128*, 892–903. https://doi.org/10.1037/abn0000446

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

Chen, J., & Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*, 759–771. https://doi.org/10.1093/biomet/asn034

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*, 6. https://doi.org/10.1186/s12864-019-6413-7

Christensen, A. P., & Golino, H. (2020). Estimating factors with psychometric networks: A Monte Carlo simulation comparing community detection algorithms. *PsyArXiv*. https://doi.org/10.31234/osf.io/hz89e

Christensen, A. P., Golino, H., & Silvia, P. J. (2020). A psychometric network perspective on the validity and validation of personality trait questionnaires. *European Journal of Personality*. https://doi.org/10.1002/per.2265

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

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis* (2nd ed.). New York, NY: Psychology Press.

Cramer, A. O. J. (2012). Why the item "23+ 1" is not in a depression questionnaire: Validity from a network perspective. *Measurement: Interdisciplinary Research & Perspective*, *10*, 50–54. https://doi.org/10.1080/15366367.2012.681973

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

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

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

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

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

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

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., Kruis, J., & Marsman, M. (2017a). Estimating psychopathological networks: Be careful what you wish for. *PLoS ONE*, *12*. https://doi.org/10.1371/journal.pone.0179891

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017b). 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. (2018b). 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

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

Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *PsyArXiv*. https://doi.org/10.31234/osf.io/zg84s

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

Friedman, J., Hastie, T., & Tibshirani, R. (2014). *glasso: Graphical lasso – estimation of Gaussian graphical models*. Retrieved from https://CRAN.R-project.org/package=glasso

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

Golino, H., & Christensen, A. P. (2020). *EGAnet: Exploratory Graph Analysis– a framework for estimating the number of dimensions in multivariate data using network psychometrics*. Retrieved from https://CRAN.R-project.org/package=EGAnet

Golino, H., Christensen, A. P., Moulder, R., Kim, S., & Boker, S. M. (2020). Modeling latent topics in social media using Dynamic Exploratory Graph Analysis: The case of the right-wing and left-wing trolls in the 2016 US elections. *PsyArXiv*. `https://doi.org/10.31234/osf.io/tfs7c`

Golino, H., & 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., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., . . . Martinez-Molina, A. (2020). Investigating the performance of Exploratory Graph Analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, *25*, 292–320. `https://doi.org/10.1037/met0000255`

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, *18*, 277–296.

Guyon, H., Falissard, B., & Kop, J.-L. (2017). Modeling psychological attributes in psychology–an epistemological discussion: Network analysis vs. latent variables. *Frontiers in Psychology*, *8*, 798. `https://doi.org/10.3389/fpsyg.2017.00798`

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. *Multivariate Behavioral Research*. `https://doi.org/10.1080/00273171.2019.1640103`

Haslbeck, J. M. B., & Waldorp, L. J. (2020). mgm: Structure estimation for time-varying mixed graphical models in high-dimensional data. *Journal of Statistical Software*, *93*, 1–46. `https://doi.org/10.18637/jss.v093.i08`

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer Science & Business Media. Retrieved from `https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12.pdf`

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.

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L. J., . . . 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`

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*, 8577–8582. `https://doi.org/10.1073/pnas.0601602103`

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`

Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, *2*, 37–63. `https://doi.org/10.1.1.214.9232`

R Core Team. (2020). *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/%2010.18637/jss.v48.i02`

Schmittmann, V. D., Cramer, A. O. J., 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`

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. `https://doi.org/10.1177/0956797611417632`

Sobel, M. E. (1997). Measurement, causation and local independence in latent variable models. In M. Berkane (Ed.), *Latent variable modeling and applications to causality* (pp. 11–28). New York, NY: Springer. `https://doi.org/10.1007/978-1-4612-1842-5_2`

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

van Bork, R., Rhemtulla, M., Waldorp, L. J., Kruis, J., Rezvanifar, S., & Borsboom, D. (2019). Latent variable models and networks: Statistical equivalence and testability. *Multivariate Behavioral Research*, 1–24. https://doi.org/10.1080/00273171.2019.1672515

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (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

Waldorp, L., & Marsman, M. (2020). Relations between networks, regression, partial correlation, and latent variable model. *arXiv*. Retrieved from https://arxiv.org/abs/2007.10656

Ward, J. H. (1963). Hierarchical clustering to optimise an objective function. *Journal of the American Statistical Association*, *58*, 238–244.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*, 440–442. https://doi.org/10.1038/30918

Yin, J., & Li, H. (2011). A sparse conditional gaussian graphical model for analysis of genetical genomics data. *The Annals of Applied Statistics*, *5*, 2630–2650.

Supplementary Materials

SI 1. Breakdown of Conditions for Simulation 2

This pattern further extended into the effects of condition where both EFA and network loadings were similarly affected by the number of variables ($\eta_p^2 = 0.09$ and 0.11, respectively) and correlations between factors ($\eta_p^2 = 0.05$ and 0.02, respectively) as well as the interaction between sample size and number of factors ($\eta_p^2 = 0.02$ and 0.03, respectively), sample size and type of data ($\eta_p^2 = 0.01$ and 0.01, respectively), and number of variables and correlations between factors ($\eta_p^2 = 0.01$ and 0.04, respectively; Figure 4).
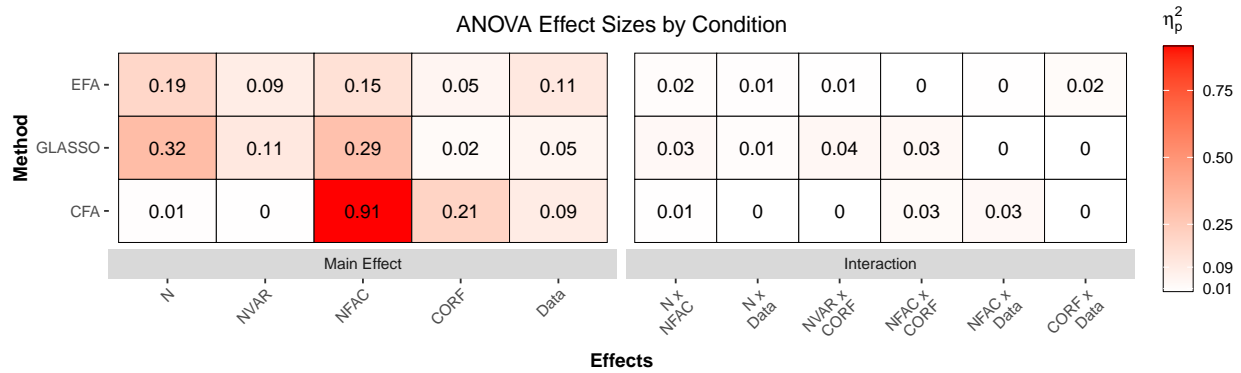

Figure 4. ANOVA partial eta-squared effect sizes broken down by main effect conditions and their interactions. Interactions that are not shown did not achieve a small effect size for any method. GLASSO = network loadings, N = sample size, NVAR = number of variables, NFAC = number of factors, CORF = correlation between factors, and Data = type of data (continuous and polytomous)

Although EFA and network loadings both had large main effects for sample size ($\eta_p^2 = 0.19$ and 0.32, respectively) and number of factors ($\eta_p^2 = 0.15$ and 0.29, respectively), the effect sizes were larger in magnitude for the network loadings. Based on Figure 2, the effect of sample size was most pronounced for the smallest sample size ($n = 250$), which had the largest mean differences in correlations in comparison to all other sample sizes ($n_{500} = 0.04$, $n_{1000} = 0.07$, and $n_{500} = 0.11$) except for the difference between $n = 500$ and $n = 5000$ (0.07; all Tukey's HSD pairwise comparison effect sizes were large). Because the

GLASSO network estimation method estimates fewer edges in smaller sample sizes, we suspect that there was a greater number of zeros in the loading matrix, which resulted in decreased accuracy. For the number of factors, the effect size for network loadings was nearly twice as large as EFA loadings. This difference was likely due to network loadings estimating more loadings near or at zero relative to EFA loadings. Indeed, the extremely large effect size for CFA loadings suggests that having zeros in the loading matrix was detrimental to correlations with the population loadings.

When looking into the interactions for network loadings, number of factors had small effects with sample size and correlation between factors (both η_p^2's $= 0.03$). The number of factors and sample size interaction is apparent in Figure 2 where accuracy tended to decrease as sample size got smaller and number of factors got larger. Similarly, the interaction between the number of factors and correlations between factors showed a decrease in accuracy as number of factors increased and correlations between factors decreased (Figure 2). This seems to be supported by the better accuracy when there was a moderate correlation (.50) between factors. Notably, there was decreased accuracy when correlation between factors was large (.70), which may have been due to the increased size of cross-loadings.

SI 2. Descriptive Simulation and Development of the Loadings Comparison Test

Methods

Random and factor models. Data for the random and factor models were generated following the same procedure as the testing simulation.

Network model. For the descriptive simulation, we kept the neighborhood of each node constant at 4 and manipulated the rewiring probability using three values 0, 0.50, and 1, which corresponded to networks that were completely regular, small-world, and completely random, respectively. It's important to note that the completely random networks were different from the random correlation matrix (some zero vs. all non-zero values, respectively).

Loadings and factors. Loadings and factors were computed following the same procedure as the testing simulation.

Statistical Analyses. For the descriptive simulation, we computed the mean and standard deviation for the proportion of network and factor loadings that were greater than or equal to their respective small, moderate, and large effect sizes. We also computed the means and standard deviations for the proportion of dominant and non-dominant loadings that were greater than or equal to small effect sizes for both network and factor loadings. These descriptive statistics were then used to formulate heuristics for an algorithm to determine the data generating model.

Results

The table below presents the descriptive statistics for the data generated from all models and conditions in the descriptive simulation. As a general finding, we observed that the network loading proportions were fairly consistent across models relative to the factor loadings. More specifically, the factor model had roughly the same proportions when the model was a factor and network model except for the dominant loadings. When the model was a random model, the factor model did not appear to differentiate the measurement of

variables on factors and had substantially low dominant loadings. In contrast, the network loadings had relatively similar proportions across models with some variation depending on the models conditions. In general, there were larger proportions of small and non-dominant loadings when the generating model was a network. Based on their relative consistency, we suggest that the network loadings may be a more general case of differentiating variables' measurement of factors.

Table 3

Descriptive Statistics for the Descriptive Simulation

Effect Size	Loadings	Random M (SD)	Factor .00	Factor .30	Factor .50	Factor .70	Network 0	Network .50	Network 1
Small	Factor	0.08 (0.02)	0.32 (0.02)	0.31 (0.02)	0.30 (0.03)	0.27 (0.06)	0.30 (0.08)	0.29 (0.08)	0.30 (0.08)
	Network	0.30 (0.15)	0.33 (0.00)	0.34 (0.01)	0.35 (0.03)	0.37 (0.06)	0.49 (0.08)	0.48 (0.10)	0.47 (0.10)
Moderate	Factor	0.05 (0.02)	0.17 (0.04)	0.16 (0.04)	0.15 (0.04)	0.12 (0.05)	0.17 (0.07)	0.16 (0.07)	0.16 (0.07)
	Network	0.08 (0.06)	0.27 (0.03)	0.23 (0.04)	0.18 (0.04)	0.12 (0.04)	0.20 (0.06)	0.17 (0.07)	0.17 (0.07)
Large	Factor	0.03 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.02 (0.03)	0.05 (0.06)	0.05 (0.06)	0.05 (0.06)
	Network	0.01 (0.02)	0.12 (0.03)	0.08 (0.03)	0.04 (0.02)	0.02 (0.02)	0.03 (0.03)	0.02 (0.02)	0.02 (0.02)
Dominant	Factor	0.25 (0.09)	0.96 (0.05)	0.94 (0.06)	0.90 (0.08)	0.78 (0.12)	0.76 (0.14)	0.77 (0.14)	0.78 (0.13)
	Network	0.64 (0.16)	1.00 (0.01)	0.99 (0.02)	0.95 (0.06)	0.80 (0.12)	0.94 (0.08)	0.91 (0.09)	0.91 (0.09)
Non-Dominant	Factor	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.01 (0.03)	0.02 (0.03)	0.02 (0.03)
	Network	0.12 (0.11)	0.00 (0.00)	0.01 (0.02)	0.07 (0.05)	0.16 (0.06)	0.21 (0.09)	0.23 (0.10)	0.23 (0.10)

Note. One thousand samples were generated per condition. Conditions are listed under their respective models (correlations between factors for Factor and rewiring probability for Network). Loadings correspond to factor and network loadings. Both dominant and non-dominant effect sizes are greater than or equal to the small effect size for the respective loading type. The mean and standard deviation values (in parentheses) correspond to the mean and standard deviation of the proportion of loadings greater than or equal to the effect size.

Random vs. non-random model. In reviewing Table 3, we start with the random data generating model. There are two main differences between the factor and network

loadings, which were in line with what we theoretically suggested. First, the network loadings have a much larger average proportion of loadings that are a small effect size (or larger) than the factor loadings. Another difference is the dominant non-dominant (or cross-) loadings proportions, which were far fewer for factor loadings relative to the network loadings. In fact, this latter pattern is observed across all data generating models, which is not surprising given factor models in particular tend to minimize cross-loadings whereas network models encourage them.

In terms of what separates the random model from a non-random model (i.e., factor or network), we noticed that the moderate and large effect size proportions for the factor loading means and standard deviations are numerically close relative to the other models. This is in line with our suggestion that factor models are less likely to achieve moderate and large loadings in a random model because the common covariance between variables is minimal. This further implies that the proportion of loadings that are at least a moderate and large effect size for factor loadings should be relatively equal across most data generated from a random model.

We verified this idea by computing the proportion of times that moderate (or larger) factor loadings were equal to large (or larger) factor loadings across the samples of all models. We found that in 48.3% of the samples this was the case when the data was generated from a random model and less than 2% in any of the other models. The largest proportion was for the small-world network with a rewiring probability of 1 (1.3%), which corresponded to a random network model. This finding suggests that this proportion can be used as a heuristic to determine a random model from a non-random model.

Factor vs. network model. As expected given our Simulation 2 results, the proportions of small, moderate, large, and dominant loadings were relatively consistent between the factor and network loadings (in line with our second simulation). There was a tendency for the network loadings to have a relatively larger proportion of moderate and large loadings when the correlation between factors were orthogonal (.00) or small (.30) as

well as a larger proportion of small loadings when the correlations between factors were large (.70). Perhaps the biggest difference was observed with the non-dominant loadings where the network loadings tended to have much larger proportions as the correlation between factors increased. This difference was also observed across network models but to a larger extent.

The clearest difference between factor and network loadings for the factor and network models was the proportion of small (or larger) loadings. Although the network loadings generally had a greater proportion of small (or larger) loadings relative to the factor loadings, this difference was much greater for network models than factor models. For the factor models, this difference was largest when the correlations between factors was large (.70; 1.37 times greater). For the network models, this difference was smallest when the rewiring probability was 1 (1.57 times greater). This fits with our expectation that network models would have a greater proportion of small loadings relative to factor models because of the difference in the way the loadings are obtained (divided across dimensions vs. maximum possible common covariance, respectively). We leveraged this finding as a heuristic to determine a network model from a factor model.

Development of the Loadings Comparison Test.   Using the heuristics we derived from the descriptive simulation, we developed an algorithm to determine whether data was generated from a random, factor, or network model. Our algorithm begins by performing a parametric bootstrap: we obtain the data and compute a covariance matrix, which we use to generate a random multivariate normal sample. Using this replicate sample, we estimate a network model using the EGA approach along with the number of factors using EGA's default Walktrap community detection algorithm. Absolute network loadings are then estimated using this information and the small, moderate, large, dominant, and non-dominant loading proportions are computed. Next, an EFA model is estimated using Maximum Likelihood and oblimin rotation with the same number of factors that was estimated by EGA. Similarly, absolute factor loadings are obtained and the small, moderate, large, dominant, and non-dominant loading proportions are computed. This process is

repeated 100 times.

After the bootstrap is completed, we then compute the proportion of samples where the moderate (or larger) and large (or larger) factor loading proportions are equal. If this proportion is greater than or equal to .20 (or 20% of the samples), then the data are predicted to be generated from a random model. If not, then the data is predicted to be from a non-random model—that is, either factor or network. To determine which model is more likely, we compute the ratio of the proportion of small (or larger) network loadings and small (or larger) factor loadings, which we refer to as the *small loading ratio.* If this small loading ratio is larger than 1.5, then the data are predicted to be generated from a network model. If not, then the data are predicted to be generated from a factor model.

In piloting the LCT on a number of different data generating structures, we observed that the the small loading ratio did not hold when network models were very sparse} (fewer than 40% of non-zero edges). Although most networks in psychology are expected to be dense, many network structures are assumed to be sparse} (Epskamp et al., 2017a). To accommodate for these sparser network structures, we added an additional heuristic to the algorithm. If the small loading ratio is less than 1.5, then the ratio between the mean dominant and mean non-dominant factor loadings across the replicate samples is computed. When reviewing} Table 3, this ratio will generally be extremely large (or infinite), but it will be much more so when the data is generated from a factor model. Taking the natural log of this ratio, we determined an appropriate cut-off of 5 or greater to predict the data are generated from a factor model; otherwise, the data are predicted to be generated from a network model. This algorithm was used as the final version of the LCT, which we have implemented in the *EGAnet* package.