UNIQUE VARIABLE ANALYSIS: A NOVEL APPROACH FOR DETECTING REDUNDANT VARIABLES IN MULTIVARIATE DATA

PSYARXIV PREPRINT

Alexander P. Christensen
Department of Neurology
University of Pennsylvania
Philadelphia, PA, 19104
alexpaulchristensen@gmail.com

Luis Eduardo Garrido
Department of Psychology
Pontificia Universidad Católica Madre y Maestra
Santiago de Los Caballeros, Dominican Republic, 51000
luisgarrido@pucmm.edu.do

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

December 29, 2020

This is a preprint of a manuscript under review.

This is the initial submission of this manuscript meaning it has not received peer-review feedback previously.

Read with skepticism. Use the Sagan standard: "extraordinary claims require extraordinary evidence"

Abstract

One common approach for constructing tests that measure a single attribute is the semantic similarity approach where items vary slightly in their wording and content. Despite being an effective strategy for ensuring high internal consistency, the information in tests may become redundant or worse confound the interpretation of the test scores. With the advent of network models, where tests represent a complex system and components (usually items) represent causally autonomous features, redundant variables may have inadvertent effects on the interpretation of their metrics. These issues motivated the development of a novel approach called *Unique Variable Analysis* (UVA), which detects redundant variables in multivariate data. The goal of UVA is to statistically identify potential redundancies in multivariate data so that researchers can make decisions about how best to handle them. Using a Monte Carlo simulation approach, we generated multivariate data with redundancies that were based on examples of known real-world redundancies. We then demonstrate the effects that redundancy can have on the accurate estimation of dimensions. Next, we evaluated UVA's ability to detect redundant variables in the simulated data. Based on these results, we provide a tutorial for how to apply UVA to real-world data. Our example data demonstrate that redundant variables create inaccurate estimates of dimensional structure but after applying UVA, the expected structure can be recovered. In sum, our study suggests that redundancy can have substantial effects on validity if left unchecked and that redundancy assessment should be integrated into standard validation practices.

***K*eywords** redundancy · minor factors · validity

Introduction

One of the core tenets of psychometrics is that a test should measure a single attribute. A primary strategy for developing tests that measure a single attribute is to select variables that are highly related to one another. This is reflected in traditional scale development conventions where variables are selected based on high item–test correlations and their contribution to the test's internal consistency (or the extent to which variables are interrelated; DeVellis, 2017). Indeed, internal consistency or Cronbach's α (Cronbach, 1951) is the most widely applied measure for the structural validation of tests, and is often used (incorrectly) to determine whether a test measures a single attribute (Flake, Pek, & Hehman, 2017).[1]

The emphasis on internal consistency has led many to recommend dropping items with low internal consistency as a common step in scale development (Boateng, Neilands, Frongillo, Melgar-Quiñonez, & Young, 2018; MacKenzie, Podsakoff, & Podsakoff, 2011). On the one hand, selecting variables with high internal consistency increases the likelihood that a single attribute is being measured (DeVellis, 2017; McDonald, 1999). On the other hand, it can lead to redundancy between some variables in the test. Incidentally, Cronbach's α is sensitive to redundancy and will tend to overestimate internal consistency in these conditions (Green & Yang, 2009).

We define *redundancy* broadly meaning the potential for a shared substantive cause (e.g., single attribute), shared semantic reference (e.g., similar item phrasing, similar item content; Rosenbusch, Wanders, & Pit, 2020), and general response tendency (e.g., social desirability; see Leising et al., 2020 for other potential causes). In classical and modern test theory, redundancy is generally thought to be beneficial for test construction because it increases test homogeneity and minimizes idiosyncratic causes (i.e., reduces measurement error; Lord & Novick, 1968). This view of redundancy is often biased towards a shared substantive cause and discounts other potential causes.

Although most instructional validation texts suggest sampling all content that is useful or relevant to the target attribute (Clark & Watson, 1995; DeVellis, 2017; McDonald, 1999), much of the literature on test development and validation suggest using the variables that lead to high internal consistency estimates for each unidimensional test (Boateng, Neilands, Frongillo, Melgar-Quiñonez, & Young, 2018; MacKenzie, Podsakoff, & Podsakoff, 2011). This recommendation is not explicit but is part of the scale development and validation process in which variables with low internal consistency are removed from the final pool of variables (Raykov, 2008). Indeed, internal consistency measures are the most widely used method for variable selection (Clark & Watson, 1995). Comrey (1988) was one advocate for using similar variables for the identification of each unidimensional factor via his *factored homogeneous item dimensions*. To construct these dimensions, Comrey (1988) suggested that the development of four-items scales should be summed up to form a *"homogeneous scale"*, implicitly suggesting the use of redundant items.

More recent perspectives on psychometrics view redundancy less favorably particularly in areas where attributes are broad such as personality (Christensen, Golino, & Silvia, 2020; Cooper, 2019; McCrae & Mõttus, 2019; Saucier & Iurino, 2020). One catalyst in this perspective shift has been the introduction of psychometric network models (Epskamp, Maris, Waldrop, & Borsboom, 2018). Psychometric network models represent a test as an interconnected network (Cramer, 2012). In these networks, nodes (circles) represent variables and edges (lines) represent conditionally dependent relations (partial correlations) between the nodes. From this representation, network proponents advocate that attributes are represented as a complex system where components of these systems are defined as *causally autonomous* or having distinct causal processes (Cramer et al., 2012). Such a perspective naturally suggests that *network components*—variables or sets of variables that share a unique common cause—should *not* be redundant (Christensen, Golino, & Silvia, 2020). Beyond theoretical reasons, redundancy can have consequences for the interpretations of psychometric networks.

[1] We note that there are other internal consistency measures such as McDonald's ω (McDonald, 1999) and other variants (McNeish, 2018) which are often better choices than Cronbach's α (especially when there are minor factors). We refer mainly to Cronbach's α in our discussion because it is still the most widely used internal consistency measure.

Potential Problems of Redundancy

A recent simulation study illuminated one psychometric consequence that redundancy can have on the interpretation of network measures. Hallquist, Wright, and Molenaar (2019) performed a set of three simulation studies to evaluate the effects of multiple latent causes on commonly used network measures called *centrality*. Centrality measures quantify the relative position of a node in a network. The three most frequently used measures have been *betweenness* (nodes used most often on the shortest path between other nodes), *closeness* (nodes with the shortest path on average to other nodes), and *strength* (absolute sum of a node's connections). In their simulations, Hallquist and colleagues (2019) generated data from common factors models, which allowed them to simulate multiple latent causes on specific variables. This approach enabled them to evaluate the effect that multiple latent causes may have on the psychometric interpretation of centrality measures.

Their most relevant simulation generated data with two factors and ten variables per factor. One variable on each factor, which we refer to as *target variables*, also loaded onto a third factor. In practical terms, such a factor might represent shared semantic similarities between two variables (e.g., "I like to go to parties" and "I like to go to art museums"; Leising et al., 2020; Rosenbusch, Wanders, & Pit, 2020). They incrementally manipulated the loadings of the target variables between the shared and non-shared factors, forming a gradient of shared to non-shared factor effects on the centrality measures. In short, they found that all measures were nonlinearly affected by the presence of multiple causes on a variable, which made centrality difficult to interpret. Because of this finding, Hallquist et al. (2019) recommended that "new measures be developed using methods that capture [causally distinct] components while discouraging the presence of latent confounding" (p. 19). In other words, redundancy broadly should be reduced in psychometric network models to more accurately interpret centrality measures.

Network measures may not be all that is affected. Dimensionality, for instance, is likely to be affected if there are strong enough relations between subsets of variables to form *minor factors* or factors of correlated residual variance (or unique variance; Garrido et al., 2020). One example comes from our own work which examined the dimensional structure of the Multidimensional Schizotypy Scale (MSS; Kwapil, Gross, Silvia, Raulin, & Barrantes-Vidal, 2018) and Multidimensional Schizotypy Scale-Brief (MSS-B; Gross, Kwapil, Raulin, Silvia, & Barrantes-Vidal, 2018) using exploratory graph analysis (EGA; Golino & Epskamp, 2017; Golino et al., 2020), a network psychometrics approach for factor estimation. The MSS and MSS-B have three theoretical factors—disorganized, negative, and positive schizotypy—that were vetted and validated using thorough psychometric practices (Gross, Kwapil, Raulin, Silvia, & Barrantes-Vidal, 2018; Kwapil, Gross, Silvia, Raulin, & Barrantes-Vidal, 2018). In two large samples ($n = 6265$ and $n = 1000$), EGA found four factors for the MSS and the theoretical three for the MSS-B (Christensen, Gross, Golino, Silvia, & Kwapil, 2019). Two of the four factors in the MSS corresponded to the theoretical disorganized (25 items) and positive (26 items) schizotypy factors. The other two factors split the theoretical negative schizotypy factor into affective (8 items) and social anhedonia (18 items) factors. This unexpected separation of the negative schizotypy factor was perhaps due to minor factors of semantically similar content rather than substantive causal differences (Leising et al., 2020).

Psychometric network models are not the only models affected by redundancy though. Redundancy can have similar consequences for factor and latent variable models as well. Inherent to latent variable models is the *principle of local independence*—that is, after conditioning variables on a latent variable, they are statistically independent (McDonald, 1999). Redundancy, regardless of how or why, will often violate this principle, leading to poor fitting models (often seen in personality; Hopwood & Donnellan, 2010; Montoya & Edwards, 2020) and confounding the interpretation of test scores (Gerbing & Anderson, 1984; Goldstein, 1980). To mitigate the effects of poor fit, common practice is to simply correlate the residual variance in the model (e.g., Cole, Ciesla, & Steiger, 2007). Such an approach only side steps the issue of potential redundancy and does not address the issue of test score interpretation (Gerbing & Anderson, 1984; Goldstein, 1980).

Strict adherence to local independence is likely unavoidable in psychology but for most validation research violations of local independence go largely unchecked (Flake, Pek, & Hehman, 2017). This suggests that many psychological tests will be *essentially unidimensional* or tapping into one dominant factor with some minor factors (Slocum-Gori, Zumbo, Michalos, & Diener, 2009). When tests are essentially unidimensional, the effectiveness of common factor analytic methods vary with no one best method performing well for accurately identifying unidimensionality (Slocum-Gori & Zumbo, 2011). In a multidimensional context, essential unidimensionality can compound the problem and lead to *overfactoring* (estimating more factors than are expected). In short, factor analytic methods are affected by redundancy just as much as network models.

Present Research

The goal for the present research was threefold. First, we wanted to demonstrate how redundancy can affect dimensionality estimation using the state-of-the-art approach of EGA. Previous research has examined and demonstrated that even a few correlated residuals can reduce model fit and lead to overfactoring in structural equation models (Montoya & Edwards, 2020). To our knowledge, effects of correlated residuals has not been explored using exploratory techniques. EGA tends to be more accurate and less biased than some of the best factor analytic methods (e.g., parallel analysis; Christensen & Golino, 2020; Golino et al., 2020); therefore, any effects that redundancy has on EGA's accuracy and bias would also be expected for any other factor analytic method. To address this goal, we performed a Monte Carlo simulation study where we manipulated sample size, number of factors, number of variables per factor, correlations between factors, and number of redundant variables within each factor. To generate data with realistic redundancies, we identified known redundancies in real-world data and mimicked their characteristics in our data generation method. We then evaluated the performance of EGA by computing measures of accuracy and bias.

Our second and primary goal was to develop a novel approach for detecting redundant variables, which we refer to as *Unique Variable Analysis* (UVA). To our knowledge, most approaches for determining whether variables are redundant require estimating a factor model first to then check the correlations between the residuals. Limitations of estimating a model first is that the correlated errors will change depending on the model (e.g., composition and number of factors), conflating the evidence for redundancy between variables. Our approach uses the association structure of the empirical data. The association structure is provided either by zero-order or partial correlations or after estimating a network and computing a measure of topological overlap (Zhang & Horvath, 2005). After, a threshold or significance test is used to determine whether pairs of variables are redundant. One advantage of our approach is that it does not require the estimation of latent factors to detect redundancy. We evaluated our approach by applying three association measures and several types of significance (e.g., standard α, multiple comparison corrections) to our simulated data. Sensitivity and specificity measures were used to quantify the extent to which these association methods and types of significance combinations had false positives (predicting variable pairs were redundant when they were not), false negatives (predicting variable pairs were not redundant when they were), and overall accuracy.

Based on our simulation results, our third goal was to demonstrate how UVA can be applied to real-world data. Using personality data with a theoretically grounded structure (e.g., five-factor model) and known real-world redundancies (e.g., "Dislike being the center of attention" and "Hate being the center of attention"), we show that redundancy can produce minor factors that lead to inaccurate estimation of the expected structure. We then provide a tutorial on how to apply UVA to this data. Afterwards, we re-estimate the dimensionality of the data and show that the expected dimensional structure can be recovered.

Methods

Data Generation

Our data generation procedure generally followed the same approach as Golino et al. (2020) and Christensen and Golino (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 with a skew of zero onto a 5-point Likert scale, mimicking many commonly used assessment instruments. We note that a skew of zero was necessary to optimally calibrate realistic redundancies in our data, which we discuss below.

Simulation Design

Across population models, factor loadings for each variable were randomly drawn from a uniform distribution with values between .40 and .70 to mimic more realistic data conditions. Similarly, cross-loadings were generated following a random normal distribution with a mean of zero and a standard deviation of .10. This procedure follows previous simulation work described in Garcia-Garzón, Abad, and Garrido (2019). These cross-loadings represent data conditions that are more likely to be found in real-word data (Bollmann, Heene, Küchenhoff, & Bühner, 2015).

Two, three, and four factors were simulated to provide multidimensional structures that are commonly found in the psychological literature (Henson & Roberts, 2006). There were six and twelve variables per factor, which were chosen to evenly produce proportions of the redundant variables per factor: 0.000, 0.167, 0.333, and 0.500. The condition of zero variables was particularly important for estimating the consistency for which methods had false positives. Correlations between factors were manipulated to be orthogonal (0.00), small (0.30), moderate (0.50), and large (0.70). Finally, very small (250), small (500), and medium (1000) sample sizes were generated to mimic typical sample sizes in psychological psychometric work.

The simulation design allowed for a mixed factorial design: $3 \times 2 \times 4 \times 4 \times 3 \times 2$ (number of factors $\times$ variables per factor $\times$ proportion of redundant items $\times$ correlations between factors $\times$ sample size $\times$ number of responses) for a total of 576 simulated condition combinations.

Simulating Redundancy

Our approach for generating realistic redundancies was a two part process: quantifying real-world redundancies and calibrating our data generating parameters. We first obtained descriptive pairwise response pattern metrics of Synthetic Aperture Personality Assessment (SAPA; Condon, 2018; see also Revelle, Dworak, & Condon, 2020) items that demonstrated obvious and substantial overlap in either content or concepts of items (see SI 1 for details about the SAPA inventory). We identified thirty-six pairs of items (see SI 2 for all item labels and their descriptions, and SI 3 for the redundant pairings). For these items, we computed pairwise response pattern metrics that quantified the proportion of exact matches between responses (PEM; a participant responded with "3" for both items) and the mean absolute error (MAE) or difference between responses on two items. If necessary, we reverse scored items so their item valences were in the same direction. The average PEM was 0.41 (*median* = 0.40, *SD* = 0.08, *range* = 0.22–0.60) and the average MAE was 0.93 (*median* = 0.91, *SD* = 0.24, *range* = 0.54–1.62).

Our data generating approach to redundancy worked as follows. First, continuous data were generated from the data generation procedure described above. Next, the number of redundant variables per factor were manipulated a priori (i.e., proportion of redundant variables). Then, from each factor, a subset of variables that was equivalent to the proportion of redundant variables multiplied by the number of variables per factor (e.g., $0.333 \times 12 = 4$ variables) was randomly sampled without replacement. This subset is referred to as the *replace* set. After, another subset of variables within the same factor were randomly sampled with replacement (excluding the subset of items already selected in the replace set). This subset is referred to as the *copy* set.

From the copy set, a certain proportion of values in each variable were copied to "replace" the corresponding values in the same positions in the replace set. To avoid replacing values with exact (continuous) values, random noise was added. This random noise, on average, added or subtracted one standard deviation from the copied value. It's worth noting that because the copy set was sampled with replacement, it was possible for variables to be redundant with more than one variable. The same values, however, were not used to avoid increasing the number of redundant pairs beyond the intended manipulation. Finally, data were categorized into a 5-point Likert scale.

For the proportion of values that were copied to replace values in the replace set, we calibrated our data generating procedure to attempt to match realistic conditions—that is, matching the observed PEM and MAE metrics mentioned above. To do this, we generated continuous data with a redundancy proportion of 0.50, categorized it, and checked our simulated data's PEM and MAE against the observed data's PEM and MAE. This process continued—generating data and checking the PEM and MAE values—until we were satisfied with the result. Our simulated data achieved an average PEM of 0.38 (*median* = 0.37, *SD* = 0.08, *range* = 0.16–0.69) and an average MAE of 0.71 (*median* = 0.72, *SD* = 0.13, *range* = 0.32–1.24). Notably, our MAE was smaller than the observed data, which suggests that the dispersion of response values for the simulated data were not as wide as the observed data.

Dimensionality Estimation

In traditional psychometric research, some redundancy has allowed scales to be developed that are aimed at targeting specific psychological attributes (DeVellis, 2017). For psychometric network models, however, this redundancy has some potential consequences such as marring the interpretability of network measures (Hallquist, Wright, & Molenaar, 2019) and obscuring dimensionality estimates due to subsets of variables forming minor factors (e.g., Christensen, Gross, Golino, Silvia, & Kwapil, 2019).

To provide a demonstration of how redundancy affects dimensionality, we applied EGA (Golino & Epskamp, 2017; Golino et al., 2020) to our simulated data (see SI 4 for technical details). Simulation studies have consistently demonstrated that EGA is as accurate as state-of-the-art factor analytic approaches, such parallel analysis, but also tends to be less biased (i.e., underfactoring or overfactoring; Christensen & Golino, 2020; Golino et al., 2020). Importantly, redundancy was not simulated in these studies and therefore remains an important condition in determining how dimensionality estimates are affected. Regardless of the dimensionality reduction method applied, redundancy is likely to affect estimates in the same way: overfactoring due to subsets of redundant variables forming minor factors.

Approach to Detecting Redundancy

To evaluate whether variables are redundant, we developed a novel approach that uses one measure of association (e.g., zero-order correlations) and one type of significance (e.g., Bonferroni). The general approach works as follows: (1) compute an association matrix, (2) obtain the lower triangle of the matrix (to count weights once), (3) remove values that equal zero so that only non-zero values remain, (4) based on the remaining absolute values estimate their best fitting empirical distribution (e.g., gamma), and (5) obtain the parameters of the empirical distribution (e.g., shape and rate) to estimate p-values for each weight. Each weight in the association matrix corresponds to one variable pair and significant weights corresponding to redundant pairs. A simpler approach, one that does not rely on significance, is to use some threshold where weights in the association matrix greater than a certain value are redundant.

We used three different association methods: zero-order correlation, partial correlation (given all other items), and weighted topological overlap (Nowick, Gernat, Almaas, & Stubbs, 2009; Zhang & Horvath, 2005). Weighted topological overlap is a network measure that determines the extent to which items in a network "overlap" by quantifying the similarity between a pair of variables' shared connections (e.g., weights, signs, quantity; see SI 5 for more details). To estimate a network, we used the standard in the psychometric network literature: the graphical least absolute shrinkage and selection operator (GLASSO; Epskamp & Fried, 2018; Friedman, Hastie, & Tibshirani, 2008; Friedman, Hastie, & Tibshirani, 2014). In applying the GLASSO, we followed the approach used in EGA (see SI 4; Golino et al., 2020).

For the types of significance, we used the standard alpha ($\alpha = .05$; hereafter referred to as alpha), adaptive alpha (Pérez & Pericchi, 2014), Bonferroni, false discovery rate (FDRalpha; Benjamini & Hochberg, 1995), and threshold. Adaptive alpha, Bonferroni, and FDRalpha were used as multiple comparison corrections. Because there are a large number of elements in an association matrix (i.e., $p(p-1)$), a multiple comparisons correction is likely necessary. Details of these significance types can be found in the Supplementary Information (SI 6).

For the thresholds, we used values of .20 for both the partial correlations and weighted topological overlap, and .70 for the zero-order correlations. The threshold for partial correlations was based on a significance value that averaged around $p = .001$ across all sample size conditions ($\bar{p} = .0009$; largest $p = 0.004$ for a sample size of 250 with 4 factors and 12 variables per factor). The threshold for weighted topological overlap was based on previous evidence reported by Nowick, Gernat, Almaas, and Stubbs (2009), which found values greater than .30 were not obtained when genes associated with the brain were permutated fifty times for each person in their sample. We opted to lower this value to .20 based on our example data which only achieved seven values greater than .30 despite the many known redundancies (see SI 3). The threshold for zero-order correlations was based on a general rule of thumb that collinearity may be a serious problem at $\geq$ .80 (Gujarati & Porter, 2008). We opted for a lower threshold of .70 to capture pairwise relations trending toward this rule of thumb.

Statistical Analyses

Effects of Redundancy on Dimensionality

We computed one measure of accuracy and one measure of bias for our dimensionality assessment. For accuracy, we computed percent correct (PC) or the average of the number of times EGA correctly estimated the population number of factors across conditions. For bias, we computed mean bias error (MBE) or the average difference between the population number of factors across conditions. The former directly captures the general accuracy of the estimation, while the latter captures whether EGA tends to underfactor or overfactor. Based on our simulation design, we expected that EGA would consistently overfactor due to the redundancies in the data. We computed analysis of variances (ANOVAs) with interactions between all conditions. Partial eta-squared (η^2) effect sizes were used to determine whether there was a main or interaction effect. We followed Cohen's (1992) effect size guidelines: small = .01, moderate = .09, and large = 0.25.

Evaluation of Redundancy Approach

To evaluate the performance of the three redundancy association methods and five types of significance (including threshold), we used sensitivity and specificity measures. We focused on true positives (TP; detecting redundancy when there was redundancy between pairs of variables), false positives (FP; detecting redundancy when there was no redundancy between pairs of variables), and false negatives (FN; not detecting redundancy when there was redundancy between pairs of variables). We did not use measures with true negatives because this would inflate all metrics due to a high number of non-redundant pairs of variables.

We computed three measures: false discovery rate (FDR; $\frac{FP}{TP+FP}$), false negative rate (FNR; $\frac{FN}{FN+TP}$), and critical success index (CSI; $\frac{TP}{TP+FP+FN}$). FDR was used to determine the number of incorrectly estimated redundant variables versus the total number of the estimated redundant variables. This measure represents an approach's tendency to over-identify redundant variables relative to the number of actual redundant variables. FNR was used to determine the number of type II errors or the number of variables that were estimated as not redundant when they were redundant versus the total number of the estimated redundant variables. This measure represents an approach's tendency to under-identify redundant variables relative to the number of actual redundant variables. CSI was used as an overall accuracy measure, giving an equal weight to true positives as false positives and negatives. This measure represents an approach's tendency to correctly detect redundant variables, with few false positives and false negatives.

Data Analysis

We used R (R Core Team, 2020) for all of our analyses and the *papaja* package (Aust & Barth, 2020) for our manuscript preparation. The package *EGAnet* (Golino & Christensen, 2020) was used to apply EGA and UVA, *psychTools* (Revelle, 2019) was used for importing the SAPA dataset, and figures were created using *ggplot2* (Wickham, 2016) and *GGally* (Schloerke et al., 2020). The session information for all necessary R packages can be found in our Supplementary Information (SI 7). All data and R scripts can be found on the Open Science Framework: `https://osf.io/9w3jy/`.

Results

Redundancy Effects on Dimensionality Estimates

We first examined the effect sizes of the conditions and their interactions using ANOVAs. For PC, this revealed that there was an overwhelmingly large effect size for the proportion of redundant variables ($\eta_p^2 = 0.52$), small-to-moderate effect size of correlations between factors ($\eta_p^2 = 0.05$), and small effect sizes of number of factors ($\eta_p^2 = 0.03$) and number of variables per factor ($\eta_p^2 = 0.02$). Notably, there was no effect for data type (continuous and polytomous), so we collapsed across this condition.

As is clear in Figure 1, redundancy had a substantial effect on EGA's accuracy to the extent that performance dropped from 93.7% with no redundancy (0.000) to 30.7% with minimal redundancy (0.167). When there was moderate (0.333) to large (0.500) redundancy, EGA was almost never estimating the proper number of factors (11.3% and 2.3%, respectively). Such negative effects on accuracy are beyond any condition examined so far in simulations using EGA (Christensen & Golino, 2020; Golino & Epskamp, 2017; Golino et al., 2020). Other effects such as correlations between factors, number of factor, and variables per factor were on par with previous results (Christensen & Golino, 2020; Golino et al., 2020).

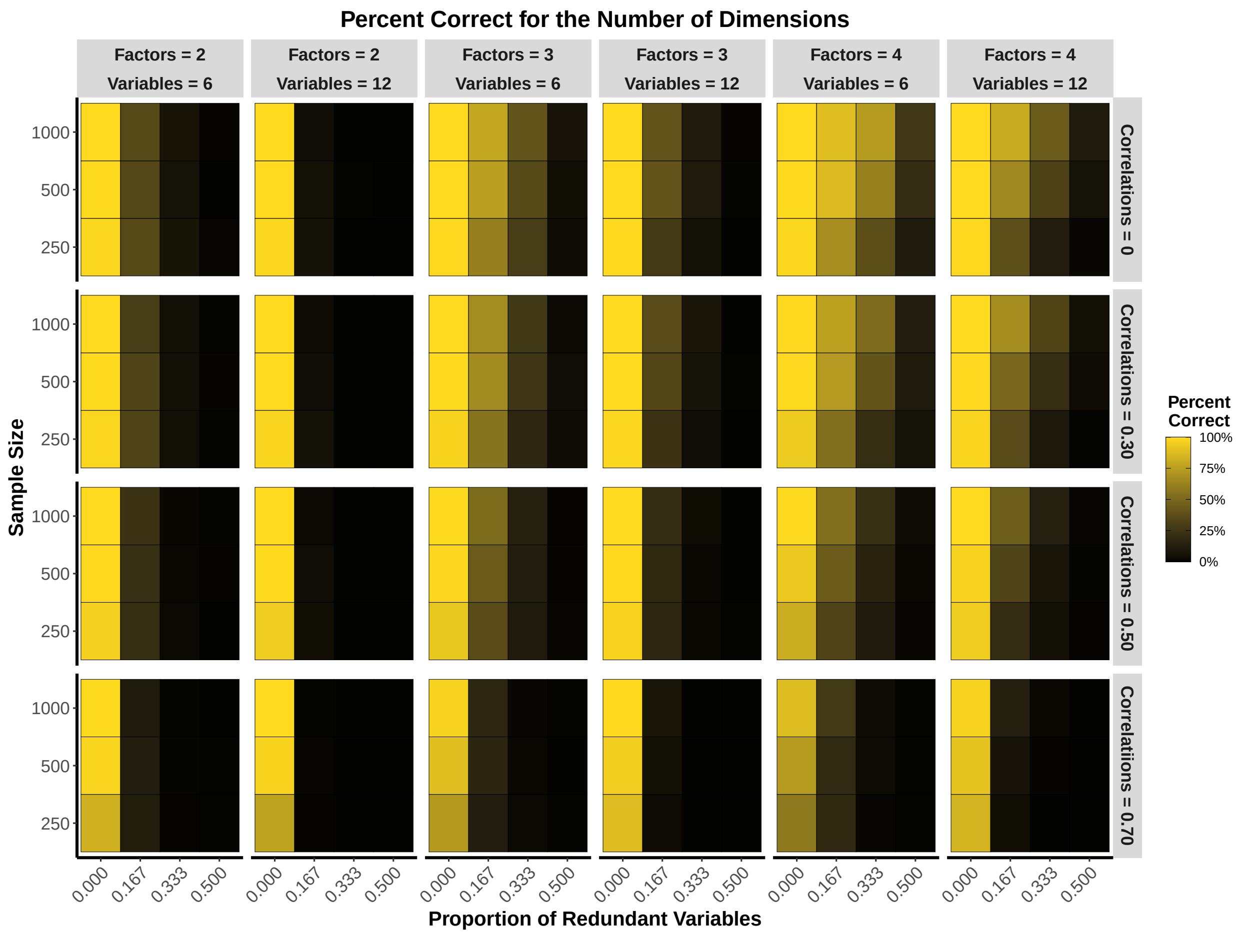

Figure 1: Redundancies Effect on the Accruacy of EGA

For MBE, there was a large main effect of proportion of redundant variables ($\eta_p^2 = 0.62$) and number of variables per factor ($\eta_p^2 = 0.23$), moderate-to-large effect of correlations between factors ($\eta_p^2 = 0.15$), and small effect of sample size and data type (both $\eta_p^2 = 0.02$). There was a moderate effect size for the interaction between proportion of redundant variables and number of variables per factor ($\eta_p^2 = 0.13$). The interaction, however, occurred below the possible value of proportion of redundant variables (0.000; see SI 8). We interpreted this interaction to suggest that as the proportion of variables increased the difference between MBE for the number of variables per factor increased with more variables having a sharper positive slope (i.e., more redundancy, the greater the increase of MBE for more variables per factors). Across all conditions, if there was bias, then it was in the direction of overfactoring (MBE $>$ 0) rather than underfactoring (MBE $<$ 0). Despite the small effect of data type, we collapsed across this condition to maintain brevity in Figure 2.

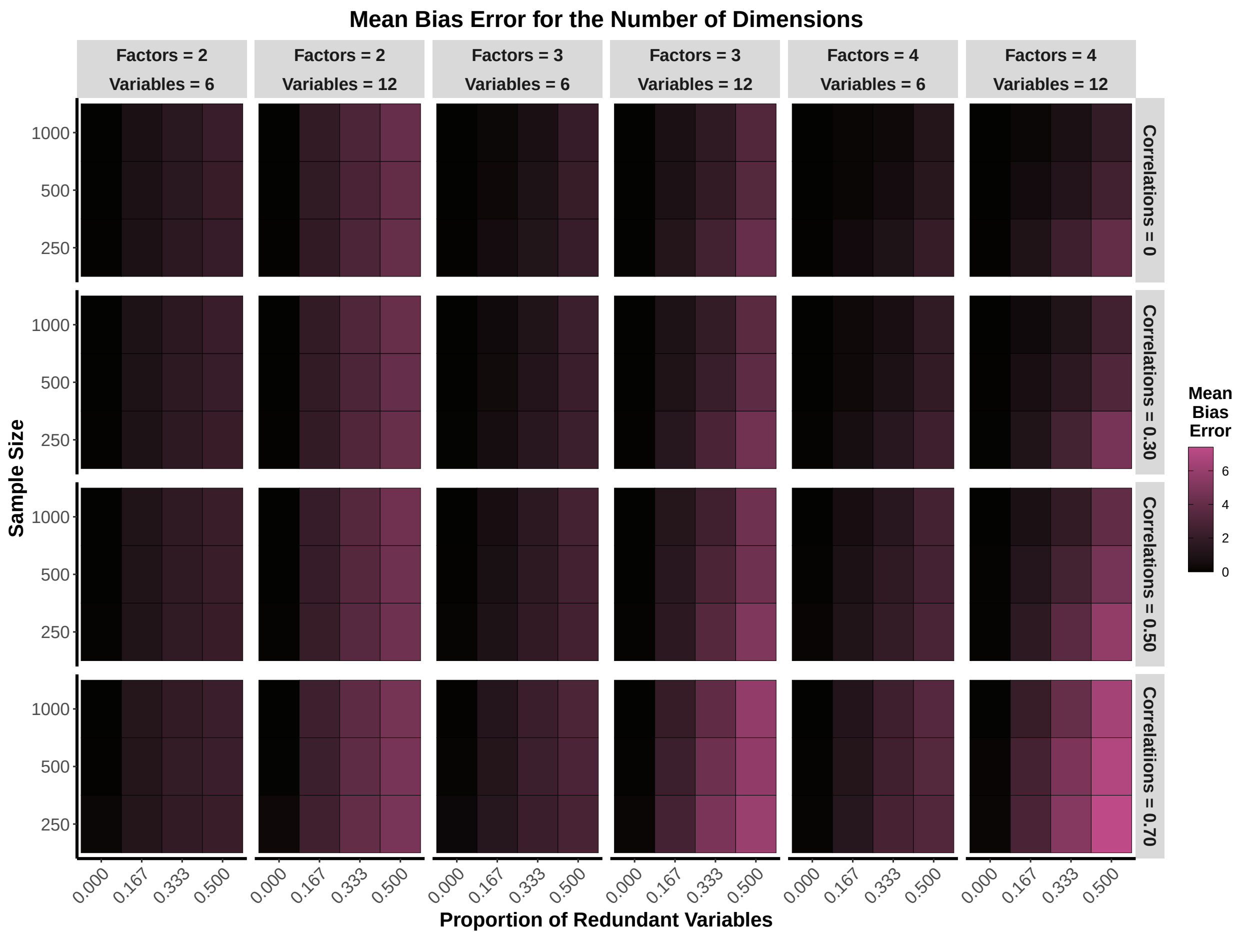

Figure 2: Redundancies Effect on the Mean Bias Error of EGA

In short, our results demonstrate that redundant variables have a strong effect on the accuracy and bias of dimensionality estimation with EGA. These effects are by far the strongest reported across all simulations that have applied EGA (Christensen & Golino, 2020; Golino & Demetriou, 2017; Golino & Epskamp, 2017; Golino et al., 2020). Moreover, redundancy interacted with number of variables per factor, suggesting that more variables and more redundancy will produce worse overfactoring bias. In short, the effects of redundancy on dimensionality estimation make clear that an approach for handling redundancies is critical.

Performance of the Redundancy Approach

No Redundancy

We start first with the results where there were no manipulated redundancies in the data. The only meaningful sensitivity and specificity measure we computed was false positives (FP). This is because there are zero true positives when there are no redundancies. For FDR ($\frac{FP}{TP+FP}$), this reduces the measure to either zero or one (whether there were any false positives); similarly, for FNR ($\frac{FN}{FN+TP}$), this reduces the measure to either zero or one (whether there were any false negatives); for CSI ($\frac{TP}{TP+FP+FN}$), this reduces the measure to always be zero because zero is the numerator.

For FP when there were zero redundancies, there was large main effect of significance type ($\eta_p^2 = 0.62$) as well as moderate-to-large main effects of number of factors ($\eta_p^2 = 0.21$) and number of variables per factor ($\eta_p^2 = 0.16$). There were also a large interaction between significance type and number of factors ($\eta_p^2 = 0.40$) as well as number of variables per factor ($\eta_p^2 = 0.39$). Moreover, there was a moderate-to-large interaction for the interaction between all three ($\eta_p^2 = 0.19$). For ease of interpretation, we present two tables with the average number of false positives for significance type and number of factors (Table 1) as well as significance type and number of variables per factor (Table 2).

Table 1: False Positives for Significance Type by Number of Factors

	Number of Factors		
Significance Type	2	3	4
Adaptive Alpha	2.40	3.78	5.49
Alpha	4.69	14.57	30.04
Bonferroni	0.00	0.01	0.01
FDR Alpha	0.01	0.02	0.03
Threshold	1.36	2.22	3.77

As is evident by the main effect of number of factors, all significance types had more FPs as number of factors increased. Bonferroni and FDRalpha were largely unaffected while alpha was substantially affected. Both adaptive alpha and threshold had more FPs as the number of factors increased, which was proportional to the number of factors increased (about two and one more FPs with each factor increase, respectively).

Table 2: False Positives for Significance Type by Number of Variables per Factor

	Number of Variables per Factor	
Significance Type	6	12
Adaptive Alpha	3.43	4.34
Alpha	6.61	26.26
Bonferroni	0.00	0.01
FDR Alpha	0.01	0.03
Threshold	2.22	2.68

Similar results held for the number of variables per factor (Table 2). Bonferroni and FDRalpha were hardly affected while alpha was substantially affected. Adaptive alpha increased one FP on average as the number of variables increased from six to twelve whereas threshold had nominal differences in FPs. Overall, these results demonstrate that FPs increase as the number of overall variables increased (i.e., number of factors $\times$ number of variables per factor). Bonferroni and FDRalpha were the best followed by threshold and adaptive alpha. Alpha, however, had a substantial number of FPs suggesting that it was ineffective at discriminating the presence of redundancy when there was none.

Redundancies

We now turn to our results for when there were manipulated redundancies. These results are presented with FDR, FNR, and CSI for minimal redundancy (0.167), moderate (0.333), and substantial redundancy (0.500). Our focus here was to compare the different association methods and significance types.

Starting with the overall accuracy (CSI) of the association method and significance type combinations, the weighted topological overlap with threshold had the highest accuracy (CSI = 0.94) followed by the weighted topological overlap with adaptive alpha (CSI = 0.90) and partial correlation with adaptive alpha (CSI = 0.88). For FDR, Bonferroni and FDRalpha were the lowest across all methods (FDR between 0.00 and 0.01) followed by weighted topological overlap with threshold (FDR = 0.06) and adaptive alpha (FDR = 0.09). For FNR, alpha and threshold with weighted topological overlap and partial correlation were the best (all FNR = 0.01). These were closely followed by adaptive alpha with weighted topological overlap and partial correlation (both FNR = 0.02). Notably, the best significance type for correlation was alpha with FNR equal to 0.05 (others ranging from 0.19–0.99). Bonferroni and FDRalpha were largely ineffective at detecting redundancies ranging from 0.67 to 0.99 across methods.

For FDR, there was a moderate main effect for proportion of redundant variables ($\eta_p^2 = 0.08$) and small main effect for number of factors for FDR ($\eta^2 = 0.03$). Because of this, we present our results by focusing on the association method, type of significance, proportion of redundant variables, and number of factors.

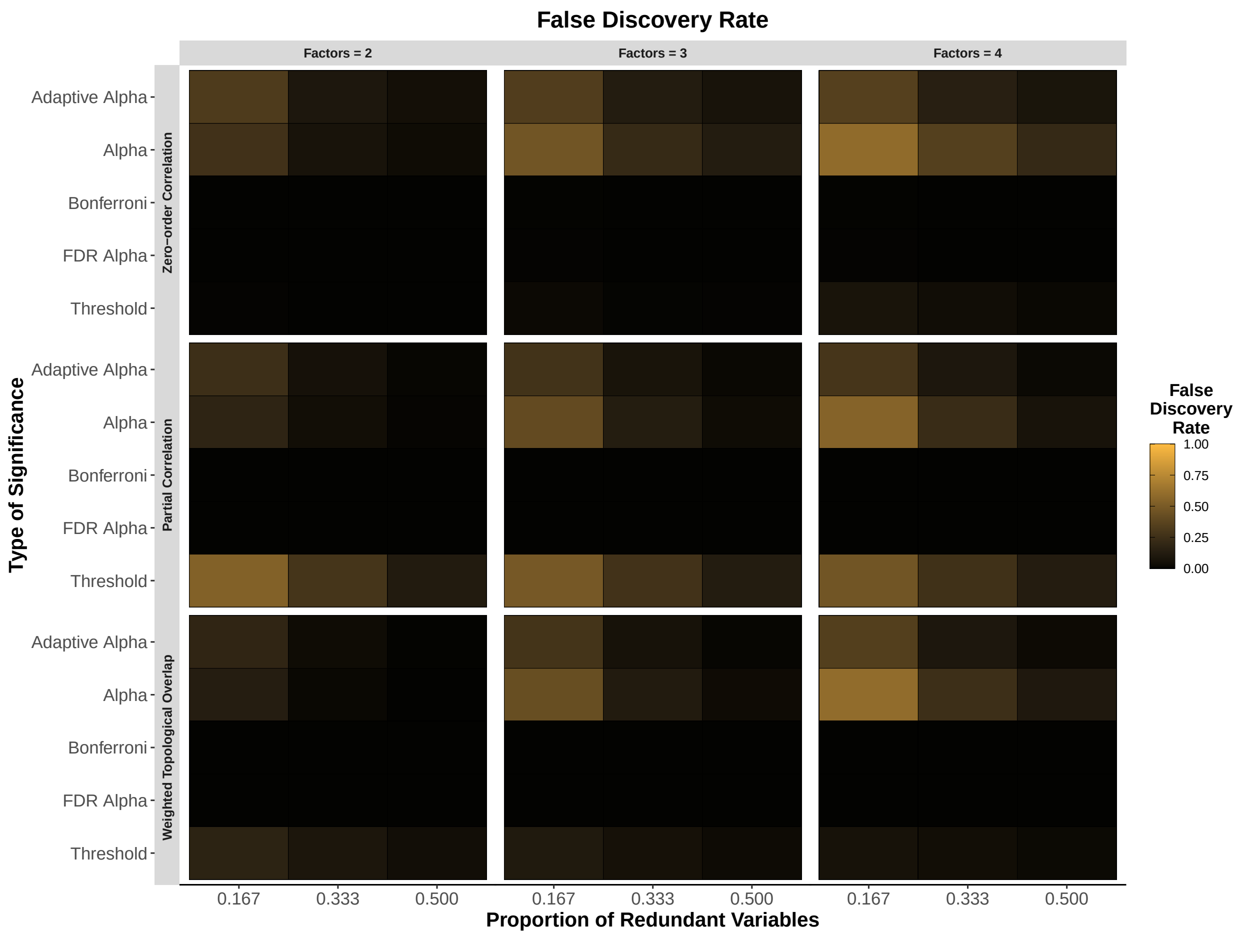

Figure 3: False Discovery Rate Across Methods and Significance

As a general trend, FDR decreased as the proportion of redundant variables increased but increased as the number of factors increased (Figure 3). Both Bonferroni and FDRalpha (FDR between 0.00 and 0.00) had few false discoveries across the association methods and the number of redundant variables, mirroring their false positives when there was no redundancy (Table 1 and 2). Threshold performed differently across the methods with low FDR for correlation (0.01) and weighted topological overlap (0.06) but moderate FDR for partial correlation (0.30). Adaptive alpha performed well across methods with low FDR (correlation = 0.10, partial correlation = 0.11, and weighted topological overlap = 0.09).

For FNR, there was only a moderate main effect for proportion of redundant variables ($\eta_p^2 = 0.18$). Relative to FDR, the opposite trend appeared in that FNR tended to increase as the proportion of redundant variables increased. This was particularly true for the Bonferroni and FDRalpha.

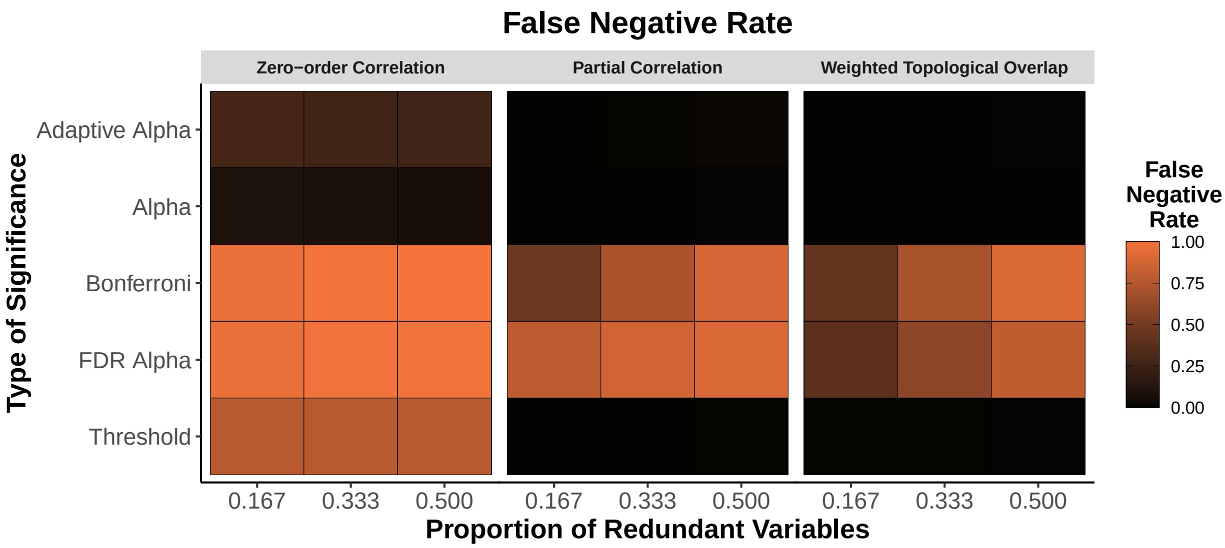

Figure 4: False Negative Rate Across Methods and Significance

For adaptive alpha, alpha, and threshold, there were practically no false negatives for the partial correlation (0.02, 0.01, and 0.00, respectively) and weighted topological overlap (0.02, 0.01, and 0.01, respectively) methods suggesting that they always found redundancies if they were there. For the correlation method, alpha had the best FNR (0.05) but this was still larger than the best performing significance types in the partial correlation and weighted topological overlap methods. Conversely, Bonferroni and FDRalpha were the worst performing, finding few redundancies (if any at all) across methods (both 0.82).

When the results of FDR and FNR are taken together, Bonferroni and FDRalpha, did not have many false positives (low FDR) but they were also not having many true positives (high FNR). This suggests that while they were not prone to false positives, they were not finding many true positives. This can be taken as evidence that these multiple correction methods were perhaps too stringent to be effective with our proposed approach. Conversely, threshold for weighted topological overlap had the some of the lowest FDR (0.06) and FNR (0.01) values across the proportion of redundant variables suggesting optimal calibration of true positive and true negative detection. Adaptive alpha followed closely behind with low FDR (0.09) and low FNR (0.02), suggesting that this combination was also detecting most redundancies when redundancies were there but at the cost of some false positives (Table 1 and 2).

For CSI, there was only a moderate main effect for proportion of redundant variables ($\eta_p^2 = 0.17$). The CSI captures FDR and FNR as an overall measure of accuracy. Figure 5 corroborates the notion that the weighted topological overlap with threshold had the best balance between FDR and FNR. Across methods, weighted topological overlap appeared to perform the best including the two most accurate significance types: threshold (0.94) and adaptive alpha (0.90). Partial correlation with adaptive alpha also performed well (0.88). The low performance of zero-order correlation across significance types (ranging from 0.01 to 0.73) suggests that zero-order correlations alone are not adequate enough to determine whether variables are redundant.

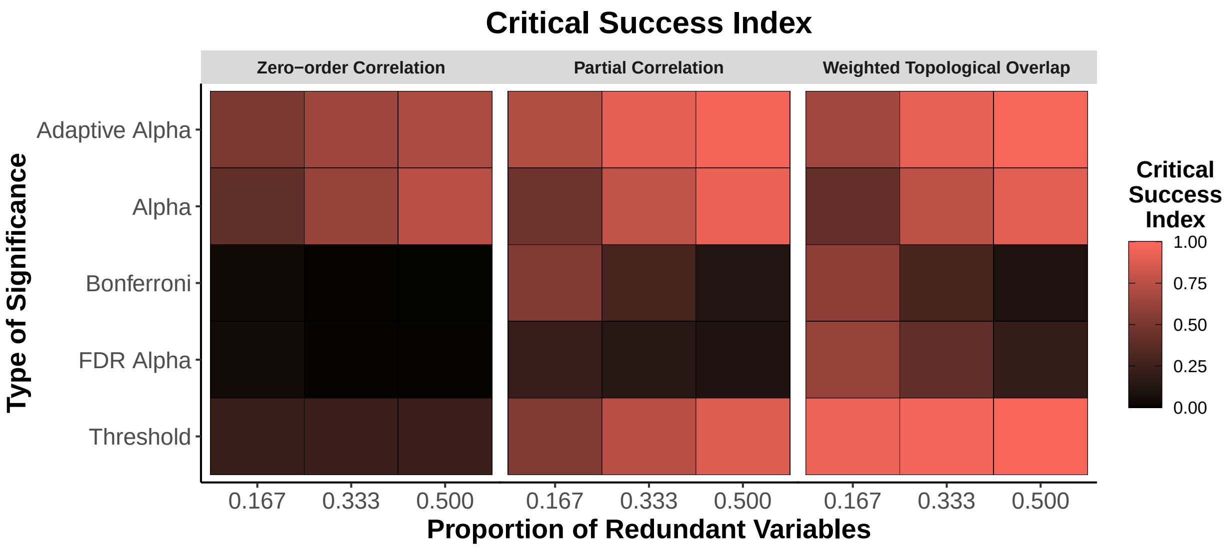

Figure 5: Critical Success Index Across Methods and Significance

Overall, weighted topological overlap with threshold performed the best and was closely followed by weighted topological overlap with adaptive alpha. Although the threshold method performed the best in our simulation, we found that it tends to err on the conservative side in real-world data. In our example data, which we turn to next, the threshold method identified "Hate being the center of attention" and "Dislike being the center of attention" as redundant as well as "Make myself the center of attention" and "Like to attract attention." It did not, however, identify the cross-combinations of the pairings—that is, "Hate being the center of attention" and "Make myself the center of attention," "Hate being the center of attention" and "Like to attract attention," "Dislike being the center of attention" and "Make myself the center of attention," and "Dislike being the center of attention" and "Like to attract attention" were *not* identified to be redundant.[2] Conversely, the adaptive alpha method identified all redundancies but perhaps also a few false positives as well. We discuss next why adaptive alpha may be more favorable.

Applied Example

For our applied example, we used the SAPA inventory (SI 1) because of its empirically validated structure of 27 components and 5 factors (Condon, 2018). We expected that our results, after applying UVA, to align with this structure—that is, the 70 items would collapse into approximately 27 unique components that have a five-factor structure. We provide an R tutorial for how to estimate factors using EGA (demonstrating overfactoring due to redundancy), apply UVA, and re-estimate factors using EGA.

Although we have demonstrated that our redundancy approach can be effective (particularly weighted topological overlap with threshold and adaptive alpha), we do not believe that UVA should be a thoughtless process. Instead, researchers should have the definitive decision on redundancy because they have theoretical knowledge about the relations between the variables. For this reason, we recommend weighted topological overlap with adaptive alpha be applied as an initial redundancy check because it tends to be slightly more liberal than threshold. From our simulation, researchers can be confident that all redundancies are detected if they're there but should use their theoretical knowledge for the final decisions. Conversely, we recommend the weighted topological overlap with threshold for adhoc checks of redundancy to determine potential redundancies that researchers might have been missed.

Based these considerations, we developed an R function in the *EGAnet* package (Golino & Christensen, 2020) called `UVA` that first applies an initial redundancy check that the researcher then verifies. After this process, an adhoc check using weighted topological overlap with threshold is performed. In our tutorial

[2]We confirmed that this result was not an artifact by reverse coding the items so that their valences were in the same direction.

below, we discuss the specific details for how to apply UVA and provide some heuristics that should guide the researcher's decision-making process through the initial redundancy check. These heuristics are not intended to be hard and fast rules but rather structured suggestions for reasoning about potential redundancies.

Getting Started

To demonstrate how to conduct UVA, we provide a tutorial of our approach applied to real-world data.[3] Before digging into UVA, the data should be loaded from the *psychTools* package (Revelle, 2019) in R.

```
# Download latest EGAnet package
devtools::install_github("hfgolino/EGAnet",
                         dependencies = c("Imports", "Suggests"))

# Load packages
library(psychTools)
library(EGAnet)

# Set seed for reproducibility
set.seed(6724)

# Load SAPA data
## Select Five Factor Model personality items only
idx <- na.omit(match(gsub("-", "", unlist(spi.keys[1:5])), colnames(spi)))
items <- spi[,idx]

# Obtain item descriptions for UVA
key.ind <- match(colnames(items), as.character(spi.dictionary$item_id))
key <- as.character(spi.dictionary$item[key.ind])
```

The code above installs the latest *EGAnet* package, loads *EGAnet* and *psychTools*, sets a seed for random number generation, and obtains the 70 SAPA items that correspond to the five-factor model of personality as well as their respective item descriptions. The item descriptions are optional but provide convenient processing when deciding which items are redundant (see Figure 7).

Handling Redundancy

Following Christensen, Golino, and Silvia (2020), `UVA` provides two approaches for reducing redundancy in data: removing all but one redundant variable or creating latent variables from redundant variables. For the former approach, researchers select one variable from variables that are determined to be redundant and remove the other variables from the dataset. As a general heuristic, researchers can compute corrected item-test correlations for the variables in the redundant response set. The variable that has the largest correlation is likely to be the one that best captures the overall essence of the redundant variables (DeVellis, 2017; McDonald, 1999). Other rules of thumb for this approach are to select variables that have the most variance (DeVellis, 2017) and variables that are more general (e.g., "I often express my opinions" is better than "I often express my opinions in meetings" because it does not imply a specific context). For the latter approach, redundant variables can be combined in to a reflective latent variable and latent scores can be estimated, replacing the redundant variables. Following recent suggestions, for ordinal data with categories fewer than six, the Weighted Least Squares Mean- and Variance-adjusted (WLSMV) estimator is used; otherwise, if all categories are greater than or equal to six then Maximum Likelihood with Robust standard errors (MLR) is used (Rhemtulla, Brosseau-Liard, & Savalei, 2012). We strongly recommend the latent variable approach because it minimizes measurement error and retains all possible information available in the data.

[3] Parts of this demonstration were taken from the Supplementary Information of our published article (Christensen, Golino, & Silvia, 2020) with some editing. The R functions used previously have changed as well as much of the information presented.

Demonstration

Initial Dimensionality Estimation

Moving forward with the application of the `UVA`, we start by evaluating the dimensional structure of the SAPA inventory *without* reducing redundancy. The following code can be run:

```
# EGA (with redundancy)
ega.wr <- EGA(items, algorithm = "louvain", plot.EGA = FALSE)
plot(ega.wr, plot.args = list(node.size = 8,
                              edge.alpha = 0.2))
```

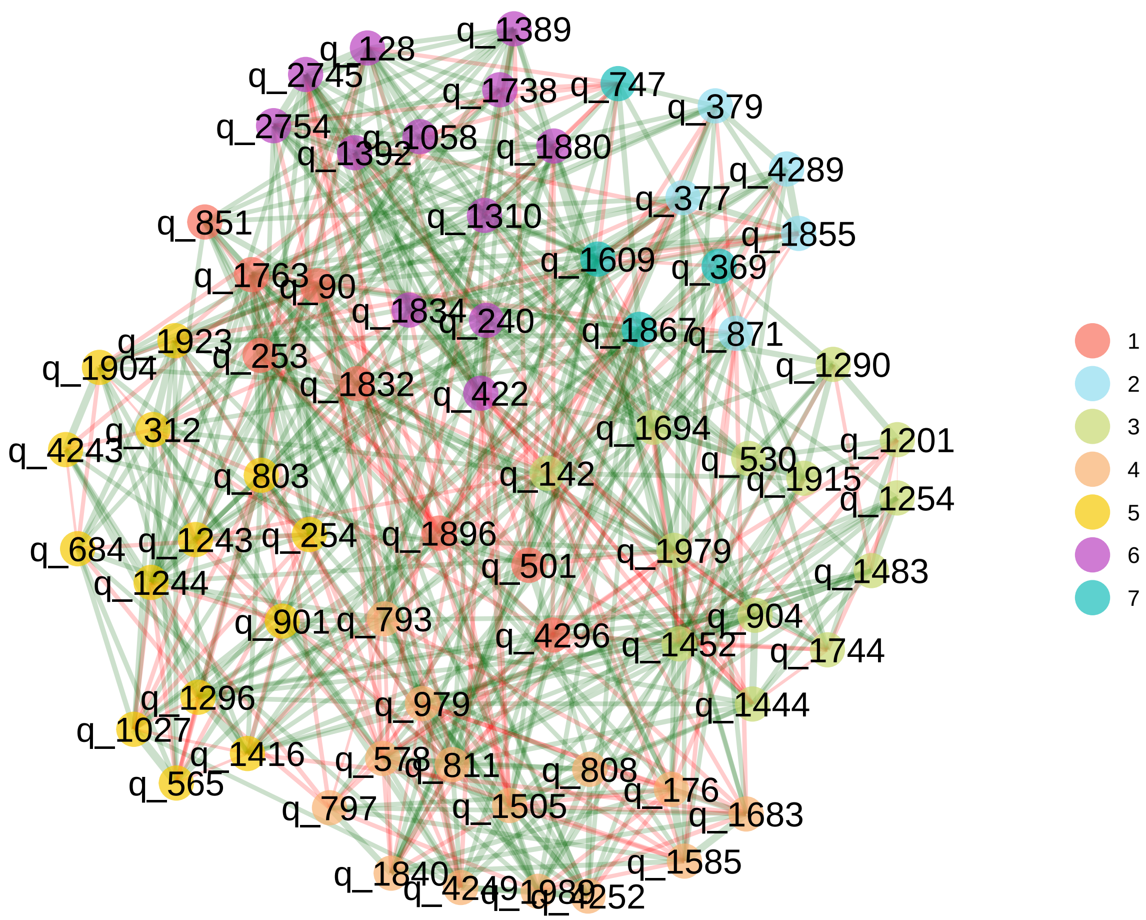

Figure 6: Exploratory Graph Analaysis of SAPA Inventory Before Unique Variable Analysis

Without performing UVA, EGA estimates that there are seven factors. Notably, there were a couple small factors identified by EGA: Factor 2 and Factor 7 (see Figure 6). Investigating the items' descriptions of these two factors, it seems likely that these represent minor factors of redundant variables: Factor 2 ("Believe that people are basically moral," "Believe that others have good intentions," "Trust people to mainly tell the truth," "Trust what people say," and "Feel that most people can't be trusted") and Factor 7 ("Enjoy being thought of as a normal mainstream person," "Rebel against authority," "Believe that laws should be strictly enforced," and "Try to follow the rules"). The divergence from the traditional five factor structure is likely due to these (and other) redundancies.[4]

UVA Tutorial

To handle the redundancy in the scale, we can now use the `UVA` function:

```r
# Perform unique variable analysis (latent variable)
sapa.ra <- UVA(data = items, method = "wTO",
               type = "adapt", key = key,
               reduce = TRUE, reduce.method = "latent",
               adhoc = TRUE)
```

There are a few arguments worth noting. First, `method` will change the association method being used. By default, the weighted topological overlap method (`"wTO"`) is applied. Second, `type` will change the significance type being used. By default, adaptive alpha is used (`"adapt"`). The `key` argument will accept item descriptions that map to the variables in the `data` argument. The `reduce` argument, which defaults to `TRUE`, is for whether the reduction process should occur. The `reduce.method` is whether the reduction process should be to create latent variables of redundant items (`"latent"`) or remove all but one of the redundant items (`"remove"`). `reduce.method` defaults to `"latent"`. Finally, `adhoc` will perform an adhoc redundancy check using the weighted topological overlap method with threshold. This check is to determine whether redundancies still might exist in the data.

Next, we'll walk through the reduction process. After running the code above, the R console will output a target variable with a list of potential redundant variables (Figure 7) and an associated "redundancy chain" plot (see Figure 8).

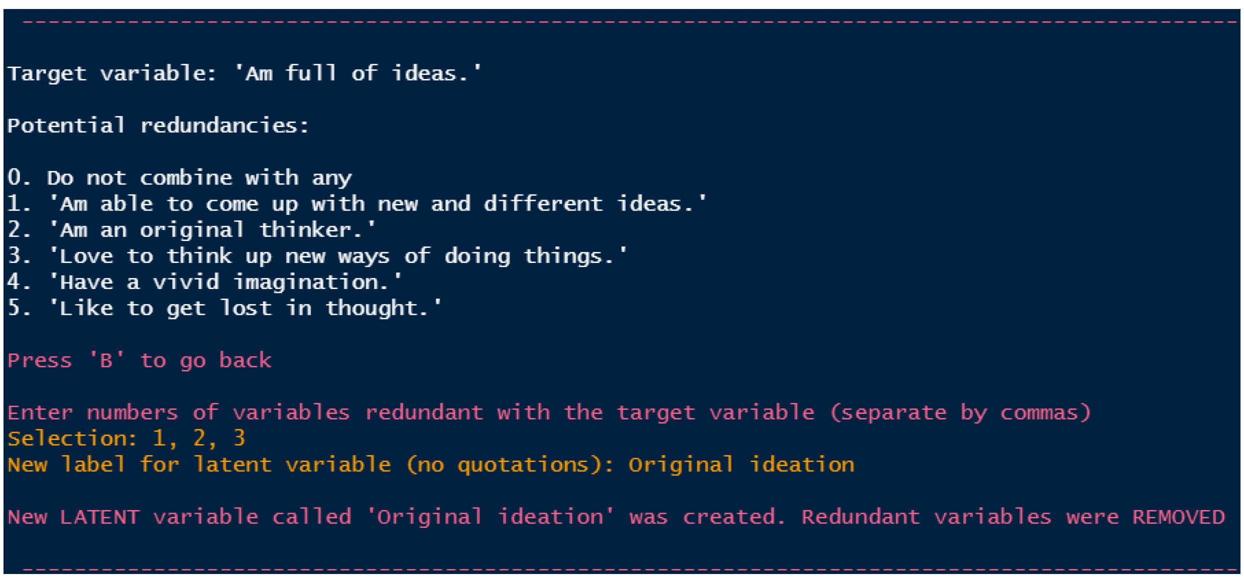

Figure 7: R Console Interface for Selecting Redundant Variables

In Figure 7, the potential redundant variables are listed below the target variable. Some of the potential redundant variables were *directly* identified as redundant with the target variable while other potential redundant variables were *indirectly* redundant meaning that they were redundant with one (or more) of the variables that were directly identified as redundant with the target variable but they themselves not actually identified as redundant with the target variable. In this way, there is a so-called "redundancy chain." Figure 8 provides a more intuitive depiction of this notion.

[4]For a comparison, we estimated dimensions using parallel analysis with polychoric correlations and principal component analysis (PCA) and principal axis factoring (PAF). These methods identified 13 and 14 dimensions, respectively.

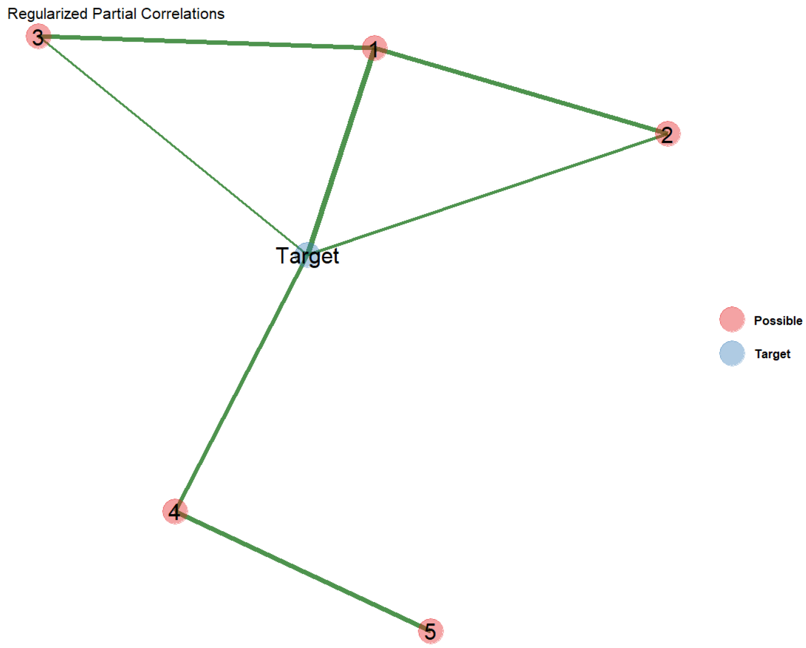

Figure 8: Example of a Redundancy Chain Plot

In the redundancy chain plot, each node represents a variable with label and color denoting the target variable ("Target" and blue, respectively) and potential redundancies (corresponding numbers and red, respectively). The connections between the nodes represent a regularized partial correlation with the thickness of an edge denoting its magnitude. The presence of an edge suggests that variables were identified as redundant rather than an actual network of associations. The interpretation of this plot would be that the target variable was identified with potential redundancy variables 1, 2, 3, and 4. Potential redundancy variable 5 was not redundant with the target variable but it was redundant with potential redundancy variable 4 (hence the "chain" of redundancy). When consulting the redundancy chain plot, researchers should pay particular attention to *cliques* or a fully connected set of nodes. In Figure 8, there are two 3-cliques (or triangles) with the target variable (i.e., Target – 1 – 2 and Target – 1 – 3).

In a typical psychometric network, these triangles contribute to a measure known as the *clustering coefficient* or the extent to which a node's neighbors are connected to each other. Based on this statistical definition, the clustering coefficient has recently been considered as a measure of redundancy in psychological networks (Costantini et al., 2019; Dinić, Wertag, Tomašević, & Sokolovska, 2020). In this same sense, these triangles suggest that these variables are likely redundant. Therefore, triangles in these redundancy chain plots can be used as a heuristic to identify redundancies.

In our example, we selected these variables as redundant by inputting their numbers into the R console with commas separating them (i.e., `1, 2, 3`). After pressing `ENTER`, a new latent variable is created from these variables and a prompt appears to label it with a new name (e.g., `'Original ideation'`). Finally, a message will appear confirming the creation of a latent variable and removal of the redundant variables from the dataset.

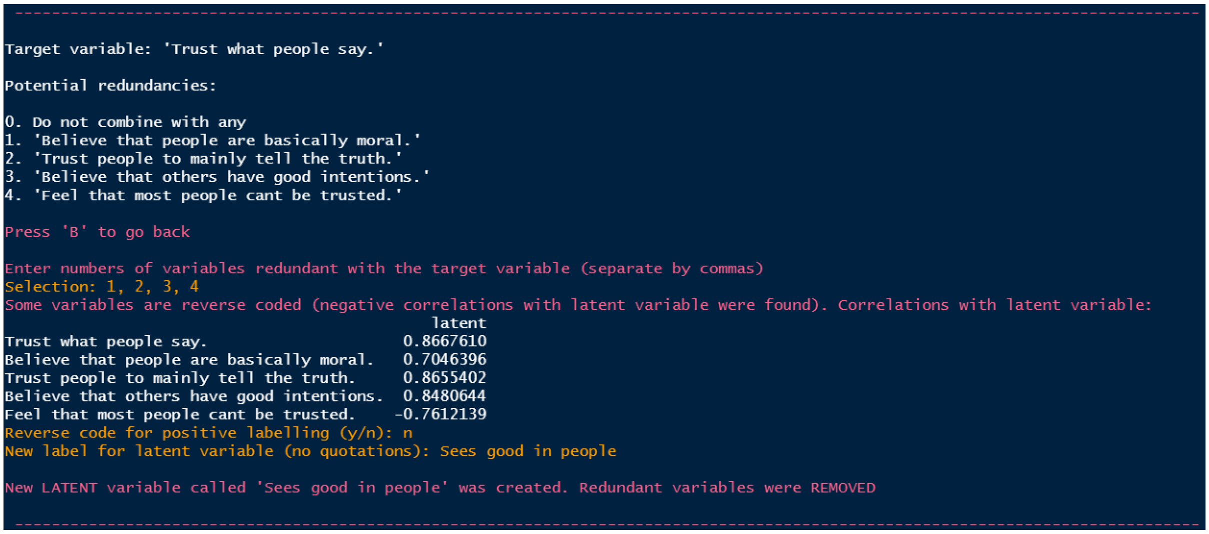

Figure 9: Second Target Variable Demonstrating Latent Variable Keying

For the second target variable (Figure 9), "Trust what people say," we combined it with all the other possible redundant items (i.e., `1, 2, 3, 4`). Notably, there was one item that was reverse keyed, "Feel that most people cant be trusted," which was negatively correlated with the latent variable. Because there was an item negatively correlated with the latent variable, a secondary prompt appears asking to reverse code the latent variable so that the label can go in the desired direction. In review of the correlations of the variables with the latent variable, we can see that the latent variable is positively keyed already; therefore, we entered `n` and labeled the component. If, however, the signs of the correlations were the inverse, then `y` could be entered, which would reverse the meaning of the latent variable towards a positively keyed orientation. The function will proceed through the rest of the redundant variables until all have been handled (see SI 9 for our handling).

Re-estimation of Dimensionality

After completing the UVA, an optional adhoc check of redundant variables can be performed using `adhoc = TRUE`. `UVA` performs this by default and will check if any redundancies remain using the weighted topological overlap method (`method = "wTO"`) and threshold (`type = "threshold"`). For our example, there were no longer any redundant variables. Our UVA reduced the dataset from 70 items down to 25 *personality components* or items or sets of items that share a common cause (Christensen, Golino, & Silvia, 2020). These components largely correspond to the 27 identified components by Condon (2018) suggesting that our approach was effective.

```r
# EGA (with redundant variables combined)
ega <- EGA(sapa.ra$reduced$data, algorithm = "louvain", plot.EGA = FALSE)
plot(ega, plot.args =
       list(vsize = 8,
            edge.alpha = 0.2,
            label.size = 4,
            legend.names = c("Conscientiousness", "Neuroticism",
                             "Extraversion", "Openness to Experience",
                             "Agreeableness")))
```

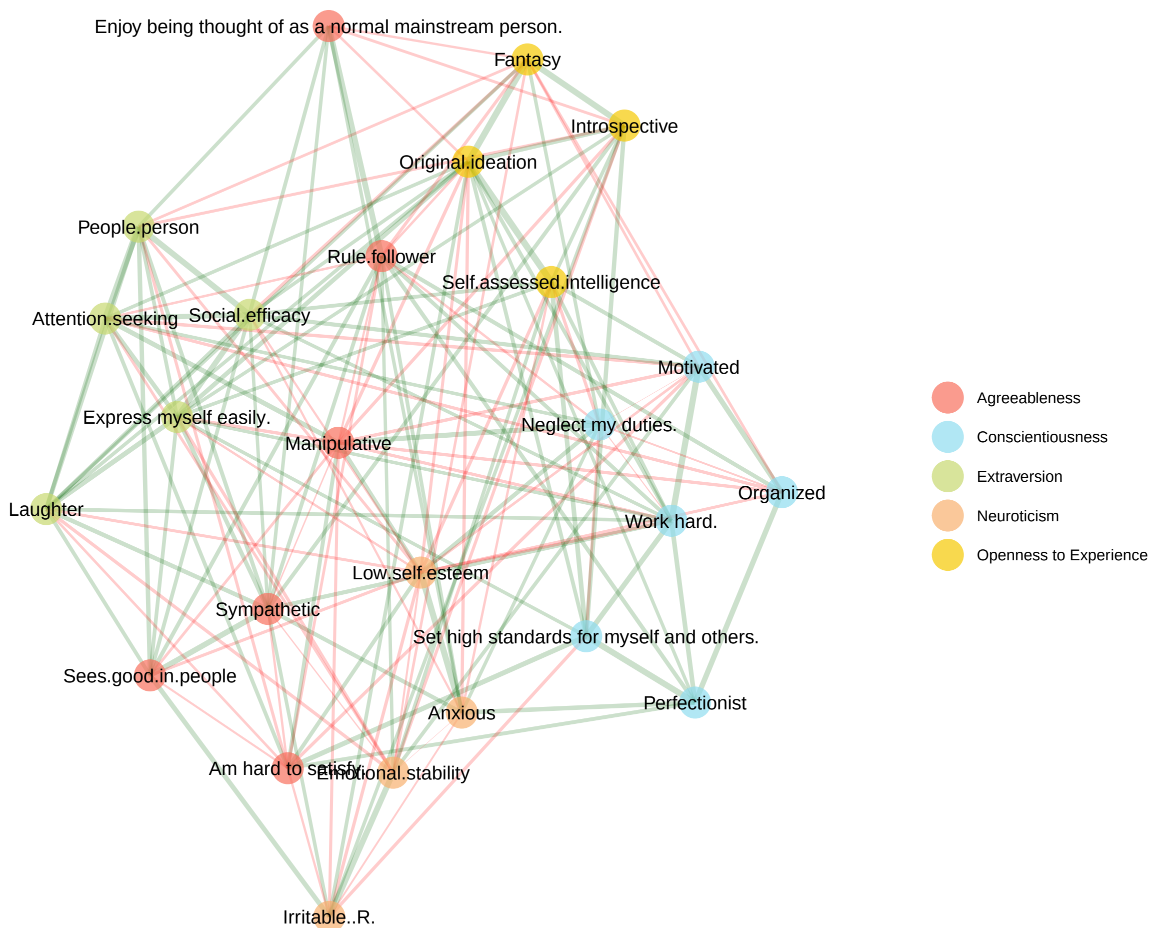

Figure 10: Exploratory Graph Analysis of SAPA Inventory After Unique Variable Analysis

With these components, we then re-estimated the dimensionality of the SAPA inventory using EGA. This time, five components resembling the five-factor model were estimated (Figure 10). These five factors also align and correspond to the expected factor structure of the SAPA inventory, corroborating the effectiveness of the UVA. In sum, our example demonstrates that redundancy can lead to minor factors, which may bias dimensionality estimates towards overfactoring (as shown in Figure 6). When this redundancy is handled, then the dimensionality estimates can be expected to be more accurate and in line with theoretical expectations (as shown in Figure 10).[5] Similar results were achieved by using the remove all but one variable approach (see SI 10).

Discussion

The present paper is the first, to our knowledge, to develop an approach for detecting redundancy (or minor factors) in multivariate data that does not require the estimation of a factor model. Our general approach relies on simple principles: use a threshold on an association matrix or create a distribution of (absolute) non-zero association weights, estimate an empirical distribution, and obtain significance values to determine which pairs of variables are redundant. The former approach is relatively inflexible to changing circumstances

[5] For a comparison, we estimated dimensions using parallel analysis with polychoric correlations and principal component analysis (PCA) and principal axis factoring (PAF). These methods identified 5 and 6 dimensions, respectively.

of the empirical data, while the latter approach almost guarantees that some false positives will occur. Despite these limitations, our simulation demonstrates that redundant variables with certain association method and significance type combinations can be detected with high accuracy. Importantly, we demonstrate that redundancy can have substantial effects on the estimation of dimensions with EGA, surpassing effects of all other conditions. Finally, we show how our novel approach, Unique Variable Analysis, can be applied to real-world data. In our empirical example, we show the effects of redundancy on dimensionality estimates and how UVA can mitigate these effects to deliver dimensionality estimates that are more in line with theoretical expectations.

The results of the dimensionality analysis in our simulation revealed strong effects of redundancy. In many regards, this result was not surprising. Our data generating method was to first generate variables with a specific factor structure and then post-hoc manipulate a certain proportion of them to be redundant. These manipulated variables were adjusted to mimic response patterns of clear and obvious redundancy in real-world data. Other variables that were not manipulated contained no redundancies other than potentially having large correlations. This means that there was a minimal gradation between non-redundant variables and variables with strong redundancy—that is, variables were either redundant or not with few opportunities to be in-between. To this extent, we created minor factors that would inevitably change the known factor structure especially with larger proportions of redundancy. Therefore, the strength of the redundancy effects on dimensionality are likely exaggerated relative to real-world data. But as Montoya and Edwards (2020) argue, it's unlikely that substantive researchers are intending to capture these minor factors (as demonstrated in our example). Nonetheless, our goal was to demonstrate the potential problems redundancy can have when left unchecked. Indeed, redundancy had the most profound effects on dimensionality estimation relative to other conditions in our simulation as well as compared to any other conditions in simulations using EGA (Christensen & Golino, 2020; Golino et al., 2020). Importantly, redundancy interacted with many other conditions, which means that its effects are pervasive and difficult to determine. Future research should attempt to simulate redundancy (even if minimally) across all variables to better mimic real-world data conditions and evaluate the persistence of the effects found in this study.

Despite the potential limitations of our minimal gradation between non-redundant and redundant variables, it was optimal for testing our novel approach to redundancy detection. Considering the strength of the redundancies, if association methods or types of significance did not work with our approach in these conditions, then they would not be expected to work in real-world circumstances. Zero-order correlations, for example, lacked enough accuracy across any significance type to be considered useful. To our surprise, the threshold of correlations greater .70 was not effective for determining redundancy despite its prevalence as a rule-of-thumb cut-off for when variables and tests might be measuring the same thing (DeVellis, 2017; Gujarati & Porter, 2008). The general lack of success for zero-order correlations might be due to the fact that they are limited to pairwise relations only. In contrast, partial correlations consider the relationship between pairwise variables given all others, which at least takes into account the pairwise relations of variables in the context of all other variables, offering relations that are conceptually closer to correlated residuals (Epskamp, Rhemtulla, & Borsboom, 2017; Waldorp & Marsman, 2020). Such an explanation is substantiated by the success of the weighted topological overlap method, which operates not only on (regularized) partial correlations but also on the extent to which two variables share similar relations—that is, their pairwise relation but also the similarity of the relations to other variables.

For the significance types, Bonferroni and FDRalpha multiple comparison corrections were far too stringent in our approach and were unsuccessful for detecting redundancy across association methods. Alpha was largely successful when it came to detecting redundancies if they were there (low FNR) but lacked precision, averaging more than fourteen false positives across conditions, for what pairs of variables were actually redundant (high FDR). Although some false positives may be reasonable, alpha had too many false positives to be effective as a guide. Rather than simplifying a researcher's selection process, substantial numbers of false positives will lead to confusion about what variables are redundant and what variables are not. For the same reason, this is what makes adaptive alpha (combined with weighted topological overlap) our recommended significance type. Adaptive alpha had low FNR across conditions meaning that it rarely missed redundancy if it was there and relatively low FDR if there was redundancy. There may indeed be false positives but these will be far fewer which makes the researchers selection process more straightforward. By comparison, threshold performed the best in our simulation but was often too strict when it came to real-world data. The weighted topological overlap with threshold may thus be the preferred choice of researchers who aim to be more conservative in determining redundancies in their data at the cost of missing some redundant variables.

One general takeaway researchers might have from our simulation and presentation of redundancy is that all redundancy is bad. This does not represent our view. Redundancy is a fundamental part of psychometric measurement. There are strong, theoretically-driven rationales for creating redundancies in tests with none more important than measuring a target attribute with minimal measurement error (DeVellis, 2017; Lord & Novick, 1968; McDonald, 1999). In this way, redundancy in narrow attributes is encouraged. If a narrow attribute is part of a broader multidimensional context, then there may be a certain point at which redundancy may become more of a weakness than a strength. Dimensionality, as we show, is strongly affected by redundancy which may lead to lower consistency of the dimensions across studies, leading to problems in understanding the substantive interpretation of the dimensions (Gerbing & Anderson, 1984; Goldstein, 1980). One reason for unreliable dimensions may be that minor factors are appearing in some studies but not others. Some of the more consequential effects may appear when attempting to measure broad attributes such as personality traits where facets contain some redundancy but not completely—that is, some items that make up a facet are redundant while others are not (Christensen, Golino, & Silvia, 2020). Oftentimes, these issues appear because of shared semantic similarity where some items are slight variations in wording and content (Leising et al., 2020; Rosenbusch, Wanders, & Pit, 2020). Undoubtedly, semantic similarity between items will create more internally consistent scales but at the cost of introducing confounds into the validity of the measurement.

Notably, there are some approaches for evaluating the extent to which tests remain cohesive. The recaptured scale technique, for example, places a test in the context of other related tests and estimates an exploratory dimensional model to assess whether the original dimensionality of the test remains (Waller, DeYoung, & Bouchard, 2016). Such an approach can be applied to unidimensional and multidimensional tests and is a direct assessment of the test's structural robustness. One limitation of this approach is that it requires collecting data from multiple tests for all participants, which is not always feasible. Another approach, bootstrap EGA (Christensen & Golino, 2019), comes from the network literature and estimates the extent to which dimensions in a test remain homogeneous and interrelated (Christensen, Golino, & Silvia, 2020). Bootstrap EGA performs a bootstrap approach on a test and estimates EGA on each bootstrap sample. Because EGA's item allocation is deterministic, the consistency of the item allocation in each dimension can be estimated. This results in a metric called *structural consistency* or the extent to which an empirically-derived dimension is exactly recovered across the bootstrap samples. In our own experience, redundancy is one of the main contributors to structural inconsistency because minor factors appear as separate factors in some replicate samples but not others. The advantage of structural consistency is that it is conceptually similar to the recaptured scale technique but does not require obtaining additional tests (although there is no reason why other tests couldn't be included).

To this end, redundancy is neither good (as commonly portrayed) or bad (as we portray). For narrow attributes, redundancy is generally good; for broad attributes, redundancy is generally bad. For psychometric network models, redundancy requires specific attention (Christensen, Golino, & Silvia, 2020; Hallquist, Wright, & Molenaar, 2019). Our Unique Variable Analysis is designed to inform researchers about the potential redundancies in their data and is equipped with approaches to handle those redundancies (creating latent variables or removing all but one variable). One opportunity for future work is to develop a measure that quantifies the optimal level of redundancy depending on the goal of the test. Some redundancy, for example, might increase the amount of information a test has while other redundancy might not. Developing an approach that determines the level of redundancy in a given test could provide guidelines about whether a researcher's test is measuring a narrow or broad attribute allowing the researcher to quantitatively assess whether they are capturing their intended breadth.

Redundancy is a fundamental feature of psychometric testing and its usefulness depends on the researchers intent. As an example, let's take the *frustration leads to aggression* hypothesis (Nesselroade & Molenaar, 2016). Students may became frustrated when trying to solve difficult mathematical problems, and some may curse the teacher while others will walk away from the class, and still some students might slap a fellow student. Cursing, walking away, and slapping someone are, in a causal latent variable perspective, three indicators of a general behavioral tendency that we can label as *"frustration produces aggression"*. The three behaviors are distinctively unique, and good indicators of the latent tendency to became aggressive due to frustration. Their correlation may be moderate-to-high, and together these indicators might produce a factor with high internal consistency. This example shows that it is possible to estimate latent factors providing the generality needed to articulate useful lawfulness without having to rely on redundant indicators. The need to use redundant indicators, as in a shared semantic reference approach (i.e., using similar item phrasing or similar item content; Rosenbusch, Wanders, & Pit, 2020), may be a well-known item-construction strategy, but it reflects either a lack of strong theory to guide the item-construction or the development of scales with inflated internal consistency. For example, *"I like going to parties"*, *"I like attending social events"* and *"I'm very fond of gatherings and festivities"* may reflect well extraversion, but won't provide differential information regarding this personality trait, since the three items are basically identical in terms of content and simply use different synonyms for *parties*. Our approach can be used to help in the identification of redundant items with an eye towards the identification of generalizable latent constructs.

When developing tests, redundancy assessment, such as the novel approach developed here, should be integrated into other test validation practices (Flake, Pek, & Hehman, 2017). We view redundancy as the first step of structural validation (Christensen, Golino, & Silvia, 2020). Importantly, our approach allows redundancy to be determined prior to estimating factors, which mitigates confounding effects of different factor structures. Our simulation demonstrated that our novel approach is effective for detecting redundancy and our empirical example demonstrated how researchers can apply our approach to their own data. Across our paper, we've discussed and demonstrated that redundancy can have substantial unintended effects if left unchecked. There is still plenty of future work to be done to understand what other consequences might exist for redundancy in data. Our hope is that researchers will become more aware of these effects and their potential consequences.

References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

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

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*, P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. https://doi.org/10.3389/fpubh.2018.00149

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

Champely, S. (2020). *pwr: Basic functions for power analysis*. Retrieved from https://CRAN.R-project.org/package=pwr

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

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

Christensen, A. P., & 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*, *34*, 1095–1108. https://doi.org/10.1002/per.2265

Christensen, A. P., Gross, G. M., Golino, H., Silvia, P. J., & Kwapil, T. R. (2019). Exploratory Graph Analysis of the Multidimensional Schizotypy Scales. *Schizophrenia Research*, *206*, 43–51. https://doi.org/10.1016/j.schres.2018.12.018

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–319. https://doi.org/10.1037/1040-3590.7.3.309

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

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*, 381. https://doi.org/10.1037/1082-989X.12.4.381

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, *56*, 754–761. https://doi.org/10.1037/0022-006X.56.5.754

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

Cooper, C. (2019). Pitfalls of personality theory. *Personality and Individual Differences*, *151*, 109551. https://doi.org/10.1016/j.paid.2019.109551

Costantini, G., Richetin, J., Preti, E., Casini, E., Epskamp, S., & Perugini, M. (2019). Stability and variability of personality networks. A tutorial on recent developments in network psychometrics. *Personality and Individual Differences*, *136*, 68–78. https://doi.org/10.1016/j.paid.2017.06.011

Cramer, A. O. 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

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334. https://doi.org/10.1007/BF02310555

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.

Dinić, B. M., Wertag, A., Tomašević, A., & Sokolovska, V. (2020). Centrality and redundancy of the Dark Tetrad traits. *Personality and Individual Differences*, *155*, 109621. https://doi.org/10.1016/j.paid.2019.109621

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

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

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

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*, 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

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 https://papers.nips.cc/paper/4087-extended-bayesian-information-criteria-for-gaussian-graphical-models

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

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

García-Garzón, E., Abad, F. J., & Garrido, L. E. (2019). Improving bi-factor exploratory modeling: Empirical target rotation based on loading differences. *Methodology*, *15*, 45–55. https://doi.org/10.1027/1614-2241/a000163

Garrido, L. E., Barrada, J. R., Aguasvivas, J. A., Martínez-Molina, A., Arias, V. B., Golino, H. F., … Rojo-Moreno, L. (2020). Is small still beautiful for the Strengths and Difficulties Questionnaire? Novel findings using exploratory structural equation modeling. *Assessment*, *27*, 1349–1367. https://doi.org/10.1177/1073191118780461

Gates, K. M., Henry, T., Steinley, D., & Fair, D. A. (2016). A Monte Carlo evaluation of weighted community detection algorithms. *Frontiers in Neuroinformatics*, *10*, 45. https://doi.org/10.3389/fninf.2016.00045

Gerbing, D. W., & Anderson, J. C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research*, *11*, 572–580. https://doi.org/10.1086/208993

Goldstein, H. (1980). Dimensionality, bias, independence and measurement scale problems in latent trait test score models. *British Journal of Mathematical and Statistical Psychology*, *33*, 234–246. https://doi.org/10.1111/j.2044-8317.1980.tb00610.x

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

Golino, H., & 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

Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, *74*, 121–135. https://doi.org/10.1007/s11336-008-9098-4

Gross, G. M., Kwapil, T. R., Raulin, M. L., Silvia, P. J., & Barrantes-Vidal, N. (2018). The Multidimensional Schizotypy Scale-Brief: Scale development and psychometric properties. *Psychiatry Research*, *261*, 7–13. https://doi.org/10.1016/j.psychres.2017.12.033

Gujarati, D. N., & Porter, D. C. (2008). *Basic econometrics* (5th ed.). New York, NY: McGraw Hill.

Gysi, D. M., Voigt, A., de Miranda Fragoso, T., Almaas, E., & Nowick, K. (2018). wTO: An R package for computing weighted topological overlap and a consensus network with integrated visualization tool. *BMC Bioinformatics*, *19*, 392. https://doi.org/10.1186/s12859-018-2351-7

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

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*, 393–416. https://doi.org/10.1177/0013164405282485

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, *14*, 332–346. https://doi.org/10.1177/1088868310361240

Kwapil, T. R., Gross, G. M., Silvia, P. J., Raulin, M. L., & Barrantes-Vidal, N. (2018). Development and psychometric properties of the Multidimensional Schizotypy Scale: A new measure for assessing positive, negative, and disorganized schizotypy. *Schizophrenia Research*, *193*, 209–217. https://doi.org/10.1016/j.schres.2017.07.001

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

Leising, D., Burger, J., Zimmermann, J., Bäckström, M., Oltmanns, J., & Connelly, B. (2020). Why do items correlate with one another? A conceptual analysis with relevance for general factors and network models. *PsyArXiv*. https://doi.org/10.31234/osf.io/7c895

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

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*, 293–334. https://doi.org/10.2307/23044045

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

McCrae, R. R., & Mõttus, R. (2019). What personality scales measure: A new psychometrics and its implications for theory and assessment. *Current Directions in Psychological Science*. https://doi.org/10.1177/0963721419849559

McDonald, R. P. (1999). *Test theory: A unified treatment*. New York, NY: Taylor & Francis. https://doi.org/10.4324/9781410601087

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*, 412–433. https://doi.org/10.1037/met0000144

Montoya, A. K., & Edwards, M. C. (2020). The poor fit of model fit for selecting number of factors in exploratory factor analysis for scale evaluation. *Educational and Psychological Measurement*. https://doi.org/10.1177/0013164420942899

Nesselroade, J. R., & Molenaar, P. C. M. (2016). Some behavioral science measurement concerns and proposals. *Multivariate Behavioral Research*, *51*, 396–412. https://doi.org/10.1080/00273171.2015.1050481

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

Nowick, K., Gernat, T., Almaas, E., & Stubbs, L. (2009). Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proceedings of the National Academy of Sciences*, *106*, 22358–22363. https://doi.org/10.1073/pnas.0911376106

Pérez, M.-E., & Pericchi, L. R. (2014). Changing statistical significance with the amount of information: The adaptive α significance level. *Statistics & Probability Letters*, *85*, 20–24. https://doi.org/10.1016/j.spl.2013.10.018

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/

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A.-L. (2002). Hierarchical organization of modularity in metabolic networks. *Science*, *297*, 1551–1555. https://doi.org/10.1126/science.1073374

Raykov, T. (2008). Alpha if item deleted: A note on loss of criterion validity in scale development if maximizing coefficient alpha. *British Journal of Mathematical and Statistical Psychology*, *61*, 275–285. https://doi.org/10.1348/000711007X188520

Revelle, W. (2019). *psychTools: Tools to accompany the 'psych' package for psychological research*. Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psychTools

Revelle, W., Dworak, E. M., & Condon, D. M. (2020). Exploring the persome: The power of the item in understanding personality structure. *Personality and Individual Differences*, 109905. https://doi.org/10.1016/j.paid.2020.109905

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*, 354–373. https://doi.org/10.1037/a0029315

Rosenbusch, H., Wanders, F., & Pit, I. L. (2020). The Semantic Scale Network: An online tool to detect semantic overlap of psychological scales and prevent scale redundancies. *Psychological Methods*, *25*, 380–392. https://doi.org/10.1037/met0000244

Saucier, G., & Iurino, K. (2020). High-dimensionality personality structure in the natural language: Further analyses of classic sets of english-language trait-adjectives. *Journal of Personality and Social Psychology*, *119*, 1188–1219. https://doi.org/10.1037/pspp0000273

Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., … Crowley, J. (2020). *GGally: Extension to 'ggplot2'*. Retrieved from https://CRAN.R-project.org/package=GGally

Slocum-Gori, S. L., & Zumbo, B. D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social Indicators Research*, *102*, 443–461. https://doi.org/10.1007/s11205-010-9682-8

Slocum-Gori, S. L., Zumbo, B. D., Michalos, A. C., & Diener, E. (2009). A note on the dimensionality of quality of life scales: An illustration with the Satisfaction with Life Scale (SWLS). *Social Indicators Research*, *92*(3), 489–496. https://doi.org/10.1007/s11205-008-9303-y

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

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

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

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer. Retrieved from https://ggplot2-book.org/

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

Supplementary Information

Unique Variable Analysis: A Novel Approach for Detecting Redundant Variables in Multivariate Data

SI 1. SAPA Inventory

Our approach to simulate redundancy started with descriptive analyses of known redundancy in a large real-world dataset. We used the Synthetic Aperture Personality Assessment (SAPA) inventory (Condon, 2018) dataset available in the *psychTools* (Revelle, 2019) package in R. A subset of these items (n = 70) form a five-factor structure that corresponds to the Five Factor Model (FFM; McCrae & Costa, 1987). The response options ranged from 1 (very inaccurate) to 6 (very accurate).

This 70-item subset was completed by 4,000 participants over the SAPA project website (https://sapa-project.org). These participants were collected after the developmental dataset (from February 2017 to May 2017) and were the first 4000 complete cases (rather than the first 4000 participants; D. Condon, personal communication). The sample had a mean age of 26.90 (SD = 11.49, range = 11–90) and were well represented for both sex (59.5% female) and education (11.1% graduated high school, 31.8% currently in university, 22% graduated university, and 11.8% held a graduate or professional degree). Race and ethnicity demographics were not provided; however, the data was gathered via the SAPA project website allowing equal opportunity for people of all ages, genders, ethnicities, and socio-economic backgrounds as long as they had access to the internet.

One potential sampling bias for this sample was that these participants were included because they completed all 135 items meaning that participants who did not complete all 135 items during the same time period were not included (regardless of whether they stopped or unintentionally skipped an item; D. Condon, personal communication). Despite this potential for representative bias, this sample likely represents a broader and more diverse population than most other self-report research in the personality literature.

There were several reasons for choosing this dataset, but we elaborate on a few only. First, the dataset is a large, diverse sample that is open-source, making the analyses performed in this study available for replication. Second, personality and self-report inventories are perhaps the most commonly used assessment instruments across psychological research and therefore represent the vast majority of the applications that the UVA will be used for. Finally, the SAPA inventory is structured hierarchically: there are 27 empirically derived lower-order dimensions that can be further collapsed into the prototypical FFM (Condon, 2018). These lower-order dimensions contain substantial redundancy, making the dataset a good example for how UVA can be applied and the number of unique components to expect after reducing the redundancy of the inventory (i.e., around 27). In addition, although the true dimensional structure cannot be known, there is strong theoretical support that five dimensions should underlie the data.

SI 2. SAPA Item Label and Description

Item Label	Description
q_90	Am concerned about others.
q_1763	Sympathize with others feelings.
q_253	Am sensitive to the needs of others.
q_1896	Use others for my own ends.
q_851	Feel sympathy for those who are worse off than myself.
q_1832	Think of others first.
q_501	Cheat to get ahead.
q_377	Believe that others have good intentions.
q_871	Feel that most people cant be trusted.
q_1855	Trust what people say.
q_4296	Tell a lot of lies.
q_142	Am hard to satisfy.
q_379	Believe that people are basically moral.
q_4289	Trust people to mainly tell the truth.
q_1290	Like order.
q_1744	Start tasks right away.
q_1979	Work hard.
q_1452	Neglect my duties.
q_1915	Want every detail taken care of.
q_1201	Keep things tidy.
q_530	Continue until everything is perfect.
q_904	Find it difficult to get down to work.
q_1867	Try to follow the rules.
q_1694	Set high standards for myself and others.
q_369	Believe that laws should be strictly enforced.

Item Label	Description
q_1444	Need a push to get started.
q_1483	Often forget to put things back in their proper place.
q_1254	Leave a mess in my room.
q_979	Get overwhelmed by emotions.
q_4252	Am a worrier.
q_1989	Worry about things.
q_1505	Panic easily.
q_4249	Would call myself a nervous person.
q_808	Fear for the worst.
q_793	Experience my emotions intensely.
q_1840	Think that my moods dont change more than most peoples do.
q_811	Feel a sense of worthlessness or hopelessness.
q_1585	Rarely get irritated.
q_578	Dislike myself.
q_176	Am not easily annoyed.
q_797	Experience very few emotional highs and lows.
q_1683	Seldom get mad.
q_1904	Usually like to spend my free time with people.
q_4243	Like going out a lot.
q_312	Avoid company.
q_565	Dislike being the center of attention.
q_1416	Make myself the center of attention.
q_1923	Want to be left alone.
q_1027	Hate being the center of attention.
q_684	Dont like crowded events.

Item Label	**Description**
q_254	Am skilled in handling social situations.
q_1296	Like to attract attention.
q_901	Find it difficult to approach others.
q_1243	Laugh a lot.
q_803	Express myself easily.
q_1244	Laugh aloud.
q_128	Am full of ideas.
q_2745	Am able to come up with new and different ideas.
q_2754	Am an original thinker.
q_1392	Love to think up new ways of doing things.
q_1058	Have a vivid imagination.
q_240	Am quick to understand things.
q_1738	Spend time reflecting on things.
q_422	Can handle a lot of information.
q_1389	Love to reflect on things.
q_1310	Like to get lost in thought.
q_1880	Try to understand myself.
q_747	Enjoy being thought of as a normal mainstream person.
q_1609	Rebel against authority.
q_1834	Think quickly.

SI 3. Known Redundant Items in the SAPA Inventory

(R) = reverse of parent label

Attention-seeking

“Dislike being the center of attention” (R), “Hate being the center of attention” (R), “Make myself the center of attention,” and “Like to attract attention”

Orderly

“Keep things tidy,” “Leave a messing my room” (R), “Like order,” and “Often forget to put things back in their place” (R)

Worrier

“Am a worrier” and “Worry about things”

Laughter

“Laugh a lot” and “Laugh aloud”

Others-oriented

“Sympathize with others’ feelings,” “Am sensitive to the needs of others,” “Am concerned about others,” and “Think of others first”

Trusting

“Trust people to mainly tell the truth,” “Trust what people say,” and “Feel that most people can’t be trusted” (R)

Sees good in people

“Believe that others have good intentions” and “Believe that people are basically moral”

Novel thinker

“Am an original thinker” and “Love to think up new ways of doing things”

Motivated

“Start tasks right away” and “Find it difficult to get down to work” (R)

High emotionality

“Experience my emotions intensely” and “Experience very few emotional highs and lows” (R)

Anxious

"Panic easily" and "Fear for the worst"

Even-tempered

"Rarely get irritated," "Seldom get mad," and "Am not easily annoyed"

People person

"Like going out a lot," "Avoid company" (R), "Usually like to spend my free time with people," and "Want to be left alone"

Socially skilled

"Am skilled in handling social situations" and "Find it difficult to approach others" (R)

Unconventional

"Enjoy being thought of as a normal mainstream person" (R) and "Rebel against authority"

SI 4. GLASSO Estimation and Exploratory Graph Analysis

Exploratory graph analysis (EGA) is a recently developed method to estimate the number of dimensions in multivariate data using undirected network models (Golino & Epskamp, 2017; Golino et al., 2020). EGA first applies a network estimation method followed by a community detection algorithm for weighted networks (Fortunato, 2010). EGA has been shown to be as accurate or more accurate than more traditional factor analytic methods such as parallel analysis (Christensen & Golino, 2020; Golino et al., 2020). EGA was applied using the *EGAnet* package (version 0.9.8; Golino & Christensen, 2020) in R (R Core Team, 2020).

Network Estimation Method

This study applied the graphical least absolute shrinkage and selection operator (GLASSO; Friedman, Hastie, & Tibshirani, 2008; Friedman, Hastie, & Tibshirani, 2014), which estimates a Gaussian Graphical Model (GGM; Lauritzen, 1996) where nodes (circles) represent variables and edges (lines) represent the conditional dependence (or partial correlations) between nodes given all other nodes in the network. The least absolute shrinkage and selection operator (LASSO; Tibshirani, 1996) of the GLASSO is a regularization technique that reduces parameter estimates with some estimates becoming exactly zero.

The LASSO uses a parameter called lambda (λ), which controls the sparsity of the network. Lower values of λ remove fewer edges, increasing the possibility of including spurious correlations, while larger values of λ remove more edges, increasing the possibility of removing relevant edges. When $\lambda = 0$, then the estimates are equal to the ordinary least squares solution for the partial correlation matrix. In this study, the ratio of the minimum and maximum λ was set to 0.1.

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 extended Bayesian information criterion (EBIC; Chen & Chen, 2008; Epskamp & Fried, 2018). In network psychometrics literature, this approach has been termed EBICglasso and is applied using the *qgraph* package (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012) in R. The EBIC model selection uses a hyperparameter gamma (γ) to control how much it prefers simpler models (i.e., models with fewer edges; Foygel & Drton, 2010). Larger γ values lead to simpler models, while smaller γ values lead to denser models. When $\gamma = 0$, the EBIC is equal to the Bayesian information criterion. Following the EGA approach (Golino et al., 2020), if there was a disconnected node in the network, then γ was decreased by 0.25 until all nodes had at least one connection or γ reached zero.

Community Detection Algorithm

The Louvain algorithm (also referred to as Multi-level; Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) is one of the most commonly applied community detection algorithms in network science (Gates, Henry, Steinley, & Fair, 2016). The algorithm begins by randomly sorting nodes into communities with their neighbors and then uses modularity (Newman, 2006) to iteratively optimize its community partitions by exchanging nodes between communities and evaluating the change in modularity until it no longer improves. Then, the algorithm collapses the communities into latent nodes and identifies edge weights with other observed and latent nodes, which provides a multi-level structure (Gates, Henry, Steinley, & Fair, 2016). In this study, the algorithm was not used to identify hierarchical community structures in the network. The Louvain algorithm was implemented using the *igraph* package (Csardi & Nepusz, 2006) in R. It's also important to note that the algorithm implemented in igraph is deterministic; however, other implementations are not (Gates, Henry, Steinley, & Fair, 2016).

SI 5. Weighted Topological Overlap

Topological overlap measures quantify the extent to which a pair of nodes have (dis)similar connections (Ravasz, Somera, Mongru, Oltvai, & Barabási, 2002). These measures became popular in the biological sciences because of their ability to identify hierarchical organizations in metabolic and genetic networks (Nowick, Gernat, Almaas, & Stubbs, 2009; Ravasz, Somera, Mongru, Oltvai, & Barabási, 2002). The original topological overlap was derived for unweighted (binary) networks (equation from methods supplement of Ravasz et al., 2002 and amended by Zhang & Horvath, 2005):

$$\omega_{ij} = \frac{l_{ij} \ + \ a_{ij}}{\min\{k_i, k_j\} \ + \ 1 \ - \ a_{ij}},$$

where $l_{ij} = \sum_u a_{iu}a_{uj}$, $k_i = \sum_u a_{iu}$, and a_{ij} is the weighted between nodes i and j. In the psychometric network literature, k_i is often referred to as *degree* or the number of connections for node i. l_{ij} corresponds to the number of nodes that both i and j are connected to. This equation holds for weighted networks where k_i would instead reference node strength for node i. Common uses of topological overlap are to form a matrix where hierarchical clustering is then performed (e.g., Ravasz, Somera, Mongru, Oltvai, & Barabási, 2002; Zhang & Horvath, 2005). Our implementation of weighted topological overlap comes from the *wTO* package (Gysi, Voigt, de Miranda Fragoso, Almaas, & Nowick, 2018), which uses Nowick and colleagues' (2009) adjustment for signed values. For the above equation, this is simply taking the absolute values of $min k_i, k_j$ and a_{ij} in the denominator.

SI 6. Significance Types

The standard alpha simply selects all weights that have a p-value less than .05. The Bonferroni correction (also known as the familywise error rate) is the standard alpha value divided by the number of comparisons (e.g., total number of weights). The FDR controls the false positive rate of significance tests by using the expected number of false positive results (e.g., 5% with an $\alpha = .05$) to adjust for the total number of significant results. The number of false positives is controlled by a q-value, which is can be set with a slightly more liberal value (e.g., $q = .10$). The q-value suggests that rather than 10% of all tests resulting in false positives, 10% of all significant results will be false positives. Finally, the adaptive alpha adjusts the standard alpha level by accounting for a reference sample size. It's well-known that as sample size increases, the likelihood of a small effect becoming significant also increases. To account for this, Pérez and Pericchi (2014) provide the following formula:

$$\alpha_{adapt} = \frac{\alpha \, * \sqrt{n_0 \, \times \, (\log(n_0) \, + \, \chi^2_\alpha(1))}}{\sqrt{n^* \, \times \, (\log(n^*) \, + \, \chi^2_\alpha(1))}},$$

where n_0 is the reference sample size, n^* is the actual sample size, and α is the standard alpha (i.e., $\alpha = .05$). The reference sample size can be computed using a power analysis. For our purposes, this power analysis was computed using the *pwr* package (Champely, 2020) in R for a correlation with a medium effect size ($r = .30$), alpha of .05, and power of .80. This yields a reference sample size (n_0) of 84.07. The actual sample size (n^*) will be the number of weights used in the distribution.

SI 7. R Session Information

```
R version 4.0.3 (2020-10-10)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19042)

Matrix products: default

locale:
[1] LC_COLLATE=English_United States.1252
[2] LC_CTYPE=English_United States.1252
[3] LC_MONETARY=English_United States.1252
[4] LC_NUMERIC=C
[5] LC_TIME=English_United States.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] kableExtra_1.3.1  EGAnet_0.9.8      psychTools_2.0.8  papaja_0.1.0.9997

loaded via a namespace (and not attached):
 [1] lattice_0.20-41    prettyunits_1.1.1  ps_1.5.0           assertthat_0.2.1
 [5] rprojroot_2.0.2    digest_0.6.27      psych_2.0.12       R6_2.5.0
 [9] evaluate_0.14      httr_1.4.2         ggplot2_3.3.2      pillar_1.4.7
[13] rlang_0.4.9        curl_4.3           rstudioapi_0.13    callr_3.5.1
[17] rmarkdown_2.6      desc_1.2.0         labeling_0.4.2     devtools_2.3.2
[21] webshot_0.5.2      stringr_1.4.0      foreign_0.8-81     dlstats_0.1.3
[25] munsell_0.5.0      compiler_4.0.3     xfun_0.19          pkgconfig_2.0.3
[29] mnormt_2.0.2       pkgbuild_1.2.0     tmvnsim_1.0-2      htmltools_0.5.0
[33] tidyselect_1.1.0   tibble_3.0.4       bookdown_0.21      viridisLite_0.3.0
[37] fansi_0.4.1        crayon_1.3.4       dplyr_1.0.2        withr_2.3.0
[41] grid_4.0.3         nlme_3.1-151       jsonlite_1.7.2     gtable_0.3.0
[45] lifecycle_0.2.0    magrittr_2.0.1     scales_1.1.1       cli_2.2.0
[49] stringi_1.5.3      farver_2.0.3       fs_1.5.0           remotes_2.2.0
[53] testthat_3.0.1     xml2_1.3.2         ellipsis_0.3.1     generics_0.1.0
[57] vctrs_0.3.6        RColorBrewer_1.1-2 tools_4.0.3        glue_1.4.2
[61] purrr_0.3.4        processx_3.4.5     pkgload_1.1.0      parallel_4.0.3
[65] yaml_2.2.1         colorspace_2.0-0   sessioninfo_1.1.1  rvest_0.3.6
[69] memoise_1.1.0      knitr_1.30         usethis_2.0.0
```

SI 8. Interaction Plots for MBE

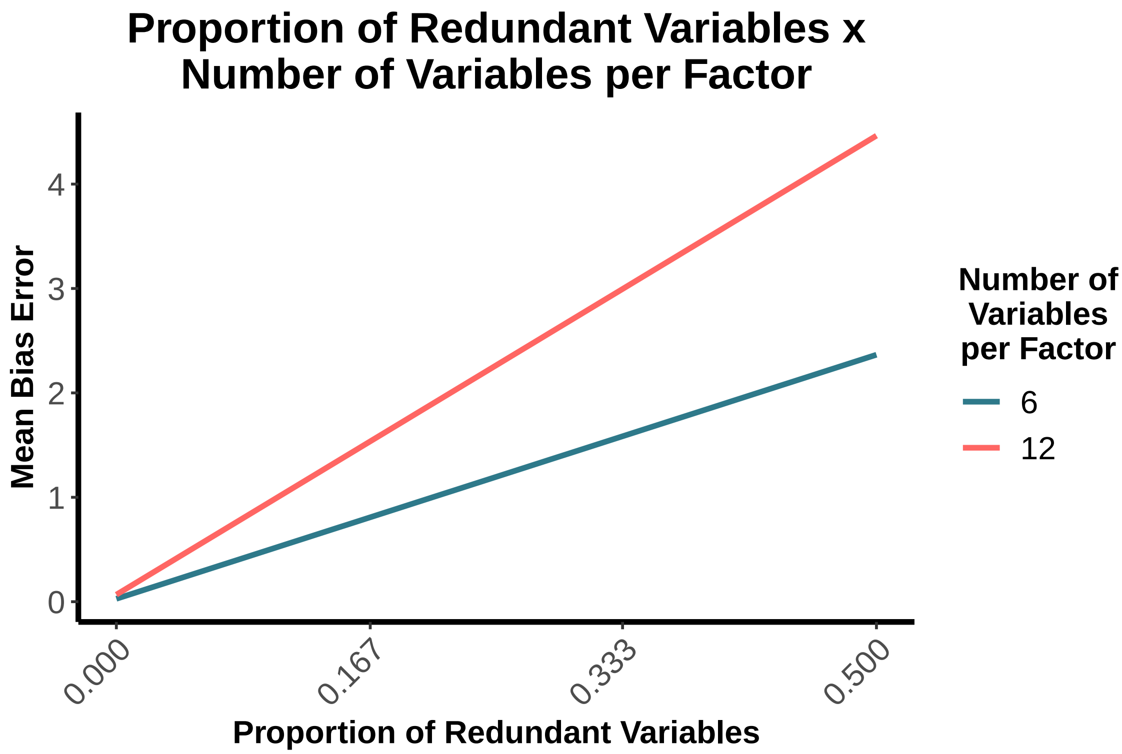

Figure 11: Mean Bias Error Interaction for Proportion of Redundant Variables and Number of Variables per Factor

10.31234/osf.io/4kra2

SI 9. Latent Variables Created During Unique Variable Analysis

		Redundant Variables			
Latent Variable	**Target**	1	2	3	4
Original ideation	Am full of ideas	Am able to come up with new and different ideas	Am an original thinker	Love to think up new ways of doing things	
Sees good in people	Trust what people say	Believe that people are basically moral	Trust people to mainly tell the truth	Believe that others have good intentions	Feel that people can't be trusted
Sympathetic	Am sensitive to the needs of others	Feel sympathy for those who are worse off than myself	Think of others first	Concerned about others	Sympathize with others feelings
Motivated	Find it difficult to get down to work	Need a push to get started	Start tasks right away		
Attention-seeking	Hate being the center of attention	Like to attract attention	Dislike being the center of attention	Make myself the center of attention	
Orgranized	Keep things tidy	Often forget to put things back in their proper place	Leave a mess in my room	Like order	
People person	Usually like to spend my free time with people	Like going out a lot	Avoid company	Want to be left alone	Don't like crowded events
Anxious	Worry about things	Would call myself a worrier	Fear for the worst	Am a worrier	Panic easily
Emotional stability	Experience very few emotional highs and lows	Get overwhelmed by emotions	Experience my emotions intensely	Think that my moods don't change more than most peoples do	
Introspective	Love to reflect on things	Try to understand myself	Spend time reflecting on things		
Irritable (reversed)	Rarely get irritated	Am not easily annoyed	Seldom get mad		
Rule-follower	Rebel against authority	Try to follow the rules	Believe that laws should be strictly enforced		
Self-assessed intelligence	Think quickly	Am quick to understand things	Can handle a lot of information		
Manipulative	Use others for my own ends	Cheat to get ahead	Tell a lot of lies		
Perfectionist	Want every detail taken care of	Continue until everything is perfect			
Low self-esteem	Feel a sense of worthlessness or hopelessness	Dislike myself			
Social-efficacy	Am skilled in handling social situations	Find it difficult to approach others			
Laughter	Laugh a lot	Laugh aloud			
Fantasy	Have a vivid imagination	Like to get lost in thought			

SI 10. Results Removing All but One Reundant Variable

In going through UVA with the remove all but one variable option (see code below), we selected the same variables as redundant as shown in SI 9 but rather than creating a latent variable we removed all but one variable.

```
# Perform unique variable analysis (removing all but one variable)
sapa.rm <- UVA(data = items, method = "wTO",
               type = "adapt", key = key,
               reduce = TRUE, reduce.method = "remove",
               adhoc = TRUE)
```

The presentation of UVA interface is mostly the same with one minor detail changed (Figure 12).

```
--------------------------------------------------------------------------------

Target variable: 'Am full of ideas.'

Potential redundancies:

0. None
1. 'Am able to come up with new and different ideas.'
2. 'Am an original thinker.'
3. 'Love to think up new ways of doing things.'
4. 'Have a vivid imagination.'
5. 'Like to get lost in thought.'

Press 'B' to go back

Enter numbers of variables redundant with the target variable (separate by commas)
Selection: 1, 2, 3

0. 'Am full of ideas.'
1. 'Am able to come up with new and different ideas.'
2. 'Am an original thinker.'
3. 'Love to think up new ways of doing things.'

Select variable to KEEP: 1

KEPT 'Am able to come up with new and different ideas.' and REMOVED all others

--------------------------------------------------------------------------------
```

Figure 12: R Console Interface for Selecting Redundant Variables

After selecting which variables are redundant, a variable is selected to be *kept*. To make this decision, the corrected item-test (or redundant set) correlations, means, standard deviations, and ranges of the variables are provided in R's plot window (Figure 13).

	Item-Total r	Mean	SD	Low	High
0 (Target)	0.76	4.78	1.15	1	6
1	0.78	4.72	1.08	1	6
2	0.68	4.56	1.13	1	6
3	0.64	4.54	1.21	1	6

Figure 13: Example of Reundancy Descriptives Table

The row names of the table denote the redundancy options which are reprinted. For the entire analysis, we selected the variables which had the largest item-test correlations (i.e., "Item-Total r" in Figure 13) and when equivalent the largest standard deviation. After UVA was finished and the adhoc check confirmed there were no more redundancies, we re-estimated the dimensionality of the dataset (Figure 14).

```
# EGA (with redundant variables removed)
ega.rm <- EGA(sapa.rm$reduced$data, algorithm = "louvain", plot.EGA = FALSE)
plot(ega.rm, plot.args = list(vsize = 8,
                              edge.alpha = 0.2,
                              label.size = 4,
                              layout.exp = 0.5,
                              legend.names = c("Conscientiousness",
                                               "Neuroticism", "Extraversion",
                                               "Openness to Experience",
                                               "Agreeableness")))
```

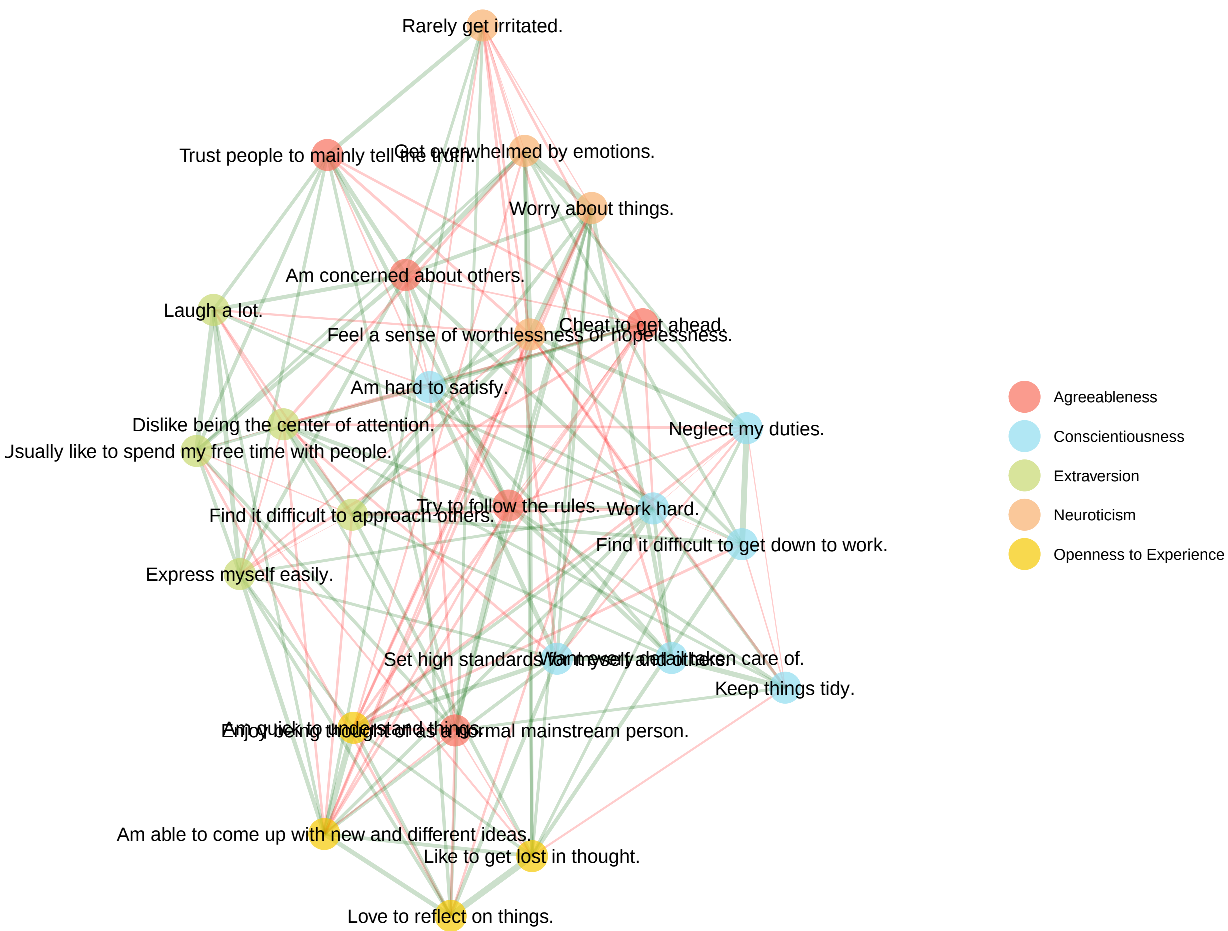

Figure 14: Exploratory Graph Analaysis of SAPA Inventory After Unique Variable Analysis (removed)

Consistent with results presented in the manuscript, five factors roughly resembling the five-factor model were found. The item placement for all items are appropriate for their dimensions as well. Similarly, parallel analysis identified five and six dimensions for principal component analysis and principal axis factoring, respectively. In all, the results largely align with one another, demonstrating that removing variables can be an effective approach to reducing redundancy in data.