



Estimating the Stability of Psychological Dimensions via Bootstrap Exploratory Graph Analysis: A Monte Carlo Simulation and Tutorial

Alexander P. Christensen^{1,*†‡}  and Hudson Golino^{2,‡} 

¹ University of Pennsylvania; alexpaulchristensen@gmail.com

² University of Virginia; hfg9s@virginia.edu

* Correspondence: alexpaulchristensen@gmail.com

† Current address: Department of Neurology, University of Pennsylvania, Philadelphia, PA, 19104

‡ These authors contributed equally to this work.

Citation: Christensen, A. P.; Golino, H. BOOTSTRAP EXPLORATORY GRAPH ANALYSIS. *Psych* **2021**, *1*, 1–23. <https://doi.org/>

Received:

Accepted:

Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Submitted to *Psych* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Exploratory Graph Analysis (EGA) has emerged as a popular approach for estimating the dimensionality of multivariate data using psychometric networks. Sampling variability, however, has made reproducibility and generalizability a key issue in network psychometrics. To address this issue, we've developed a novel bootstrap approach called Bootstrap Exploratory Graph Analysis (bootEGA). bootEGA generates a sampling distribution of EGA results where several statistics can be computed. Descriptive statistics (median, standard error, and dimension frequency) provide researchers with a general sense of the stability of their empirical EGA dimensions. Structural consistency estimates how often dimensions are replicated exactly across the bootstrap replicates. Item stability statistics provide information about whether dimensions are unstable due to misallocation (e.g., item placed in the wrong dimension), multidimensionality (e.g., item belonging to more than one dimension), and item redundancy (e.g., similar semantic content). Using a Monte Carlo simulation, we determine guidelines for acceptable item stability. After, we provide an empirical example that demonstrates how bootEGA can be used to identify structural consistency issues (including a fully reproducible R tutorial). In sum, we demonstrate that bootEGA is a robust approach for identifying the stability and robustness of dimensionality in multivariate data.

Keywords: dimensionality; network psychometrics; stability

Introduction

Exploratory Graph Analysis (EGA) [1,2] is a recently developed network psychometrics approach for identifying dimensions in multivariate data using network models. Network models are depicted by nodes (circles), which represent variables, and edges (lines), which represent the relation (e.g., partial correlation) between two nodes. *Communities* or clusters of nodes represent dimensions in networks. EGA is based on the estimation of a network followed by the application of a community detection algorithm [3].

Although EGA has become a popular approach for dimensionality assessment in the psychometric network literature, there remains lingering questions related to the stability of its results. The reproducibility of psychometric networks is an area that has received a lot of attention in recent years [4–7]. Variability, whether related to random sampling or parameter estimation, can have downstream effects on analyses that are applied to a network such as centrality measures (i.e., measures that quantify the relative position of a node in the network) and community detection algorithms. For the latter, variability

may lead to dimensional structures that do not generalize to other samples, potentially leading to inconsistencies in the number and content of the communities in the network. The goal of the present paper is to introduce a novel bootstrap approach that provides information about the stability of EGA’s results.

Exploratory Graph Analysis

The most commonly applied network estimation method in the literature is the graphical least absolute shrinkage and selection operator (GLASSO) [8,9]. The GLASSO estimates a Gaussian Graphical Model (GGM) [10] where edges are the partial correlation between two nodes given all other nodes in the network. The GLASSO penalizes and shrinks coefficients (with small coefficients going to zero) to provide more parsimonious results and lower the potential for spurious relations.

The GLASSO uses a regularization technique called the least absolute shrinkage and selection operator (LASSO) [11]. The LASSO has a hyperparameter 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 $\lambda = 0$, the estimates are equal to the ordinary least squares solution for the partial correlation matrix.

A common approach in the network psychometrics literature is to compute models across several values of λ (usually 100) and select the model that minimizes the extended Bayesian information criterion (EBIC) [12,13]. The EBIC has a hyperparameter (γ), which controls how much it prefers simpler models (i.e., models with fewer edges) [14]. 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. The EGA approach starts with $\gamma = .50$ and lowers γ by .25 if there are any unconnected nodes in the network (i.e., if a node does not have an edge to any other nodes). If γ reaches zero, then the network is used regardless of whether there are unconnected nodes in the networks.

More recently, the triangulated maximally filtered graph (TMFG) [15] has been applied as an alternative network estimation method in EGA [2,16]. The TMFG method uses a structural constraint that limits the number of zero-order correlations included in the network ($3n - 6$; where n is the number of variables). The TMFG algorithm begins by identifying four variables which have the largest sum of correlations to all other variables. Then, it iteratively adds each variable with the largest sum of three correlations to nodes already in the network until all variables have been added to the network. The end result is a network of 3- and 4-node cliques (i.e., a set of connected nodes), which form the constituent elements of an emergent hierarchy in the network [17].

In psychometric networks, clusters (or sets of many connected nodes) often emerge, forming distinct communities or sub-networks of related nodes. To estimate communities, community detection algorithms can be applied [3,18,19]. The most commonly applied algorithm used in EGA is the Walktrap community detection algorithm. The Walktrap algorithm estimates the number and content of a network’s communities using “random walks” over the network [20]. These random walks iteratively traverse over neighboring edges, with larger edge weights (i.e., partial correlations) being more probable paths of travel. Each node is repeatedly used as a starting point where *steps*—a jump from one node over an edge to another—are taken away from that node, forming a community boundary. A node’s community is then determined by its proportion of many, densely connected edges to few, sparsely connected edges (commonly optimized by a statistic known as *modularity*) [21]. The algorithm is deterministic meaning the number and variable content of the communities are estimated without the researcher’s direction. These communities are shown to be consistent with latent factors of factor models [1]. For more details about all network estimation methods and the Walktrap algorithm, see the [Materials and Methods](#) section.

83 EGA Compared to Factor Analytic Methods

84 There have been several simulation studies comparing EGA to factor analytic methods.
 85 In Golino and Epskamp’s [1] seminal simulation, EGA was compared against several
 86 commonly used factor analytic methods including the Kaiser-Guttman rule (eigenvalues
 87 > 1) [22,23], minimum average partial procedure [24], and parallel analysis (PA) [25].
 88 Overall, EGA and PA with principal axis factoring (PAF) performed the best of the
 89 methods with 96% and 97% overall accuracy (respectively). EGA tended to perform
 90 better when there fewer variables per factor (5) and moderate-to-high correlations between
 91 factors (.50 and .70), while PA tended to perform better when there were more variables
 92 per factor (10) and moderate-to-high correlations between factors.

93 The second simulation expanded on the first by adding the TMFG network estimation
 94 method and PA with principal component analysis (PCA), estimating unidimensional
 95 structures, manipulating skew, factor loadings, and data categories (continuous and
 96 dichotomous variables) [2]. Once again, EGA with GLASSO and PA with PCA and PAF
 97 had the best overall accuracy (88%, 82%, and 83%, respectively). EGA with TMFG
 98 was in the middle of pack with an overall accuracy of 73%. The PA methods tended to
 99 perform better when there were more variables per factor (8 and 12), while EGA with
 100 GLASSO tended to perform better when there were fewer variables per factor (3 and 4).
 101 EGA with TMFG performed best when there were 4 and 8 variables per factor. EGA
 102 with GLASSO (96%) and PA with PCA (100%) performed better than PA with PAF
 103 when the data were unidimensional, while EGA with GLASSO (84%) and PA with PAF
 104 (83%) performed better than PA with PCA when the data were multidimensional. EGA
 105 with TMFG trailed behind all three with 79% accuracy for unidimensional structures
 106 and 73% accuracy for multidimensional structures.

107 A more recent simulation assessed different community detection algorithms (including
 108 the Walktrap) combined with EGA with GLASSO compared to the PA methods using
 109 continuous and polytomous (5-point Likert scale) data [18]. In this study, factor loadings
 110 were drawn randomly from a uniform distribution between .40 and .70. The Louvain
 111 (90%) [26], Fast-greedy (89%) [27], and Walktrap (88%) [20] algorithms had comparable
 112 overall accuracy with PA with PCA (88%) and PAF (85%). The number of factors
 113 had the biggest effects on performance, particularly when there were more factors (e.g.,
 114 4), where the Louvain and Fast-greedy algorithms were comparable to PA with PAF
 115 (around 90%) but the Walktrap was much less accurate (around 80%). In sum, EGA
 116 with GLASSO performs comparably to common factor analytic methods like parallel
 117 analysis. Despite evidence of EGA’s effectiveness, there has yet to be an investigation
 118 into the stability of its results.

119 Bootstrap Approach

120 To address this issue, we introduce a novel approach called *Bootstrap Exploratory Graph*
 121 *Analysis* (bootEGA) to estimate the stability of dimensions identified by EGA. This
 122 approach allows for the consistency of dimensions and items to be evaluated across
 123 bootstrapped EGA results, providing information about whether the data are consistently
 124 organized in coherent dimensions or fluctuate between dimensional configurations. On
 125 the one hand, the number of dimensions identified may vary depending on the size of
 126 the sample or the sample being measured (e.g., measurement [in]variance). On the other
 127 hand, the number of dimensions might be consistent across samples but some items may
 128 be identified in one dimension in one sample and in another dimension in a different
 129 sample. Even still, some sets of items may be forming completely separate dimensions.

130 Using a bootstrap approach, the bootEGA algorithm uses one of two data generation
 131 approaches: parametric or resampling (non-parametric). The *parametric* procedure begins
 132 by estimating the empirical correlation matrix. This correlation matrix is then used as
 133 the covariance matrix to generate data from a multivariate normal distribution. The
 134 *resampling* or non-parametric procedure is implemented by resampling with replacement

from the empirical dataset. The parametric procedure should be preferred when the underlying data are expected to follow a multivariate normal distribution. The resampling procedure should be preferred when the underlying distribution is unknown or non-normal. Regardless of procedure, the same number of cases as the original dataset are used to generate or resample data. For each replicate sample, EGA is applied, resulting in a sampling distribution of EGA results. This process continues iteratively until the desired number of samples is achieved (e.g., 500).

Descriptive Statistics

From this sampling distribution, several descriptive statistics are obtained. The median number of dimensions estimated from the replicate samples is computed. The standard deviation of the number of dimensions is used as the standard error. A normal approximation confidence interval is computed by using this standard error and multiplying it by the 2.5 and 97.5 percentile of a t -distribution with iterations minus one degrees of freedom [28]. Finally, a percentile bootstrap confidence interval is computed using the same percentiles in the bootstrap quantiles [29]. In addition, a median (or typical) network structure is estimated by computing the median value of each edge across the replicate networks, resulting in a single network. Such a network represents the “typical” network structure of the sampling distribution. The community detection algorithm was then applied, resulting in dimensions that would be expected for a typical network from the EGA sampling distribution.

Dimension and Item Stability

Because EGA assigns variables to dimensions, additional metrics related to the stability of dimensions and items across bootstrap replicates can be computed. bootEGA evaluates the stability of dimensions using a measure called *structural consistency*. We define structural consistency as the proportion of times that each empirically derived dimension (i.e., the result from the initial EGA) was exactly recovered (i.e., identical item composition) from the replicate bootstrap samples was computed. Below, we provide an example of how structural consistency is computed.

Let w be the placement of items in the empirical EGA communities:

$$w = \{1, 1, 1, 2, 2, 2, 3, 3, 3, \dots, j, j, j\},$$

where values j denote dimensions (e.g., 1 denotes dimension 1). Let b^i be the bootstrap placement of items for the i^{th} iteration of bootEGA. For example, the first iteration:

$$b^1 = \{1, 1, 1, 2, 2, 2, 3, 3, 3, \dots, j, j, j\}.$$

If all elements corresponding to the j^{th} dimension of the empirical item placement and the bootstrap iteration placement match, then the dimension is given a value of 1 for the iteration; otherwise 0. For example, the first three elements of the empirical item placement w correspond to dimension 1. The first three elements of b^1 are also all 1 and therefore match (1). The next three elements of the empirical item placement w correspond to dimension 2. For b^1 , these elements are $\{2, 2, 3\}$ and therefore do *not* match (0). Finally, the last three elements of the empirical item placement w correspond to dimension 3. For b^1 , these elements are all 3 and therefore match (1).

Let d_j^i be a vector storing the values of whether the bootstrap placement of items match the empirical placement of items for dimension j and iteration i . For the example above, $d^1 = \{1, 0, 1\}$. The structural consistency of each dimension is then defined as:

$$s_j = \frac{\sum_i^n d_j^i}{n},$$

where n is the number of iterations and s_j is the structural consistency of dimension j .

Substantively, we interpret structural consistency as the extent to which a dimension is interrelated (internal consistency) and homogeneous (test homogeneity) in the presence of other related dimensions [30]. Such a measure provides an alternative yet complementary approach to internal consistency measures in the factor analytic framework.

A complementary measure to structural consistency is *item stability*. Item stability quantifies the robustness of each item's placement within each empirically derived dimension. Item stability is estimated by computing the proportion of times each item is placed in each dimension. This metric provides information about which items are leading to structural consistency (replicating often in their empirically derived dimension) or inconsistency (replicating in often in other dimensions).

One obstacle for computing item stability is that item placement is not always consistent with the empirical placement. As demonstrated in b^1 above, dimension 2 was only two elements and dimension 3 was four elements relative to both having three elements each in the empirical item placement. Below are several examples that can be compared with w :

$$\begin{aligned} w &= \{1, 1, 1, 2, 2, 3, 3, 3\} \\ b^1 &= \{1, 1, 1, 2, 2, 3, 3, 3\} \\ b^2 &= \{3, 3, 3, 1, 1, 2, 2, 2\} \\ b^3 &= \{1, 1, 1, 1, 1, 2, 2, 2\} \\ b^4 &= \{3, 3, 2, 2, 2, 1, 1, 1\}. \end{aligned} \tag{1}$$

Starting with b^2 , the composition of w and b^2 are the same but their number assignments differ. This is a common occurrence as EGA does not always assign dimension 1 to the label number 1 as it is in w . In this example, the solution is obvious: 3 becomes 1, 1 becomes 2, and 2 becomes 3.

For b^3 , the categorization is not immediately clear. First, there are only two dimension labels (1 and 2). Second, the elements corresponding to the second empirical dimension overlap in b^3 : $\{1, 1, 2\}$. To circumvent these issues, the item stability algorithm follows several steps to produce what we call ‘‘homogenized’’ item placements. First, the Adjusted Rand Index (ARI) [31] is computed between each dimension's elements of the item placement in the bootstrap iteration (b^3) and the corresponding elements for the empirical item placements (w). The ARI is computed using the following formula:

$$ARI = \frac{\binom{N}{2}(a+d) - [(a+b)(a+c) + (c+d)(b+d)]}{\binom{N}{2}^2 - [(a+b)(a+c) + (c+d)(b+d)]}$$

where a is the count of items placed into the same bootstrap and empirical community and d is the count of items that are in different communities of both the bootstrap and empirical communities. Both b and c count the wrong placement of nodes between the bootstrap community and empirical community, respectively.

For dimension 1 of b^3 and the corresponding elements in w ($\{1, 1, 1, 2, 2\}$), the ARI equals 0.4. For dimension 2 of b^3 and the corresponding elements in w ($\{2, 3, 3, 3\}$), the ARI equals 0.5. The next step orders the dimensions based on highest to lowest ARI first.

211 Ties are decided based on the number of items in the respective dimensions from highest
 212 to lowest. This ordering ensures that (1) dimensions that correspond closest to the
 213 empirically defined dimensions (e.g., w) are handled first and (2) that dimensions that
 214 are most intact (i.e., that have more items) are handled prior to equivalent ARI values.

215 The final step is identifying the mode in which the respective elements of w belong
 216 to. For dimension 3 of b^3 , the mode is 1 and therefore converts all labels of 3 to 1.
 217 For dimension 2 of b^3 , the mode is 3 and therefore converts all labels of 2 to 3. The
 218 result is a “homogenized” item placement for b^3 of $\{1, 1, 1, 1, 3, 3, 3, 3\}$. Following these
 219 steps for b^4 , the result is $\{1, 1, 2, 2, 2, 3, 3, 3\}$. Once all bootstrap item placements have
 220 been homogenized with the empirical item placements, then item stability values can be
 221 computed. These values are computed by taking the number of times an item is assigned
 222 to each dimension and dividing it by the total number of bootstrap iterations.

223 Present Research

224 In the present paper, we introduce bootEGA to estimate the stability of dimensions
 225 identified by EGA. The goal of bootEGA is to provide a tool for assessing whether
 226 dimensions and items in a scale are generalizable. Unstable dimensions and items
 227 are important for diagnosing psychometric issues that may lead to inappropriate scale
 228 interpretations. Items sorting into different dimensions, for example, may be an issue of
 229 sample size, hint that the items are multidimensional, or that they are redundant with
 230 one another and are forming a minor factor [32,33]. Therefore, an approach that is able
 231 to assess the quality of EGA’s results is necessary and would allow researchers to examine
 232 how their results would be expected to generalize to other samples. Such an approach
 233 would not only lead to more accurate interpretations but also provide researchers with
 234 greater insights into the structure of their scales.

235 To investigate the validity of the bootEGA approach and item stability analysis, we
 236 conducted a large simulation study. In this study, we manipulated sample size (250,
 237 500, and 1000), number of factors (2 and 4), number of variables per factor (4 and 8),
 238 correlations between factors (.00, .30, .50, and .70), and data categories (continuous and
 239 polytomous). Importantly, we generated data with cross-loadings and factor loadings
 240 that varied randomly between .40 and .70. We also added skew to the polytomous data
 241 ranging from -2 to 2 on 0.5 increments. The first aim was to determine whether the typical
 242 structure of bootEGA could improve the accuracy of empirical EGA’s dimensionality
 243 accuracy and item placement. The second aim was to determine guidelines for what
 244 values of item stability reflect a “stable” item. For this latter aim, we used data from the
 245 largest sample size condition only to ensure the robustness of the guidelines we propose.

246 After, we provide an empirical example to show how to apply and interpret bootEGA
 247 output. In the example, we demonstrate that bootEGA can be used to improve the struc-
 248 tural integrity of assessment instruments through the identification of problematic items.
 249 We show that problematic items can arise because several items are multidimensional
 250 (i.e., they replicate in other dimensions and cross-load onto more than one dimension).
 251 As part of the empirical example, we provide tutorial code that researchers can apply to
 252 their own data. The R code for the simulation, analyses, and example are available on
 253 our Open Science Framework repository: <https://osf.io/wxdk7/>.

254 Results

255 Dimensionality Accuracy

256 Our first goal was to determine whether the typical bootEGA network structure could
 257 improve on the dimensionality accuracy and item placement of the empirical EGA
 258 structure. Despite generating both continuous and polytomous data, there was not effect
 259 of data categories. Therefore, we collapsed across data categories. The typical structure
 260 of bootEGA was comparable to the empirical structure in terms of accuracy. The type of
 261 bootstrap did not have a substantial influence on the percent correct (PC; identifying

the same number of dimensions as the number of simulated dimensions), with all PCs within a single percent of one another for both the GLASSO and TMFG. Empirical GLASSO was slightly more accurate (87.5%) than bootEGA GLASSO (86.1%), while empirical TMFG was less accurate (79.6%) than bootEGA TMFG (82.2%). There were three main effects that were at least a moderate effect size: number of variables per factor ($\eta_p^2 = 0.06$), number of factors ($\eta_p^2 = 0.13$), and correlations between factors ($\eta_p^2 = 0.18$). There were no interactions between the conditions. Figure 1 breaks these results down by the conditions that had main effects.



Figure 1. Accuracy of the Median bootEGA Network Structure and Empirical EGA

These main effects are made clearer by Figure 1: accuracy increased as variables per factor increased, accuracy decreased as number of factors increased, and accuracy decreased as correlations between factors increased. The worst condition for all methods was when there were four variables per factor and four factors. In this condition, GLASSO remained fairly high in accuracy (around 70–90%) as correlations between factors increased until large correlations (0.70) where their accuracy dropped by about 40%. In contrast, TMFG had large decreases in accuracy with each increase in correlations between factors (around 20%). The differences between bootEGA and EGA were subtle. bootEGA tended to increase accuracy as correlations between factors increased. This pattern was particularly true for the TMFG method, suggesting that it benefits from obtain a typical structure more than the GLASSO method.

It's noteworthy that there was no effect of sample size given that number of cases tend to affect most methods [34]. Digging into our results, there was an apparent ceiling effect across conditions that might have limited the effect of sample size. Specifically, there only appeared to be performance difference when there were less favorable conditions: few variables (4) and large correlations between factors (0.70).

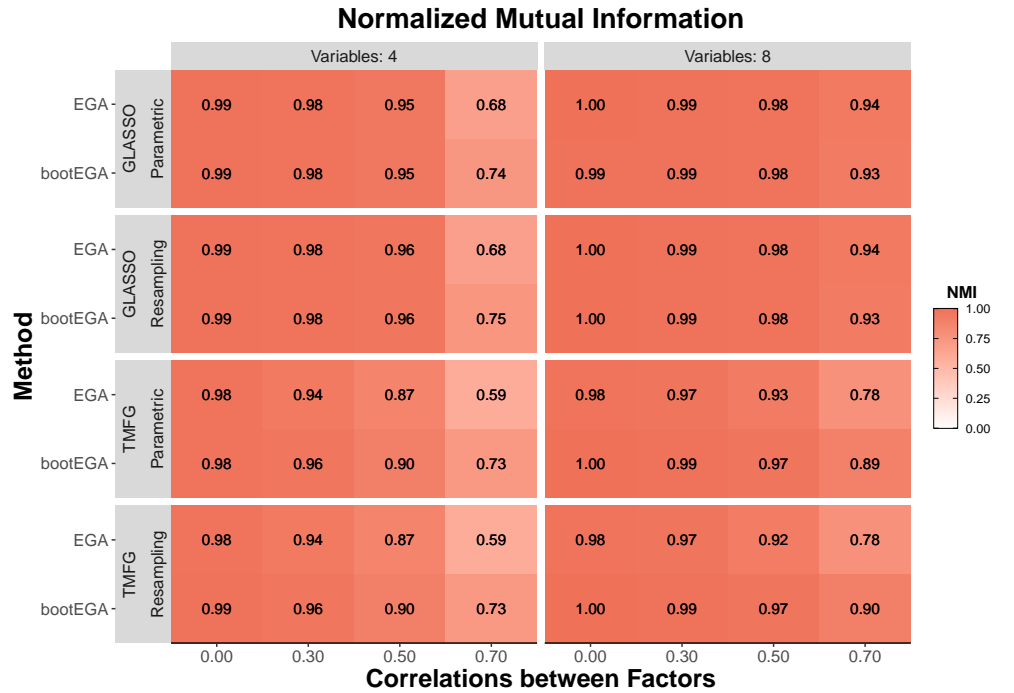


Figure 2. Normalized Mutual Information of the Median bootEGA Network Structure and Empirical EGA

Similar to accuracy, there was not an effect of data categories with item placement (normalized mutual information; NMI; see the [Materials and Methods](#) section for mathematical notation). Further, the typical structure of bootEGA was comparable to the empirical EGA. The type of bootstrap also did not have an effect on item placement. bootEGA (0.944) and empirical (0.941) EGA were comparable for GLASSO, but bootEGA (0.928) was higher than empirical (0.881) for TMFG. There was a moderate interaction between number of variables and correlations between factors ($\eta_p^2 = 0.07$).

Figure 2 shows that this interaction was mainly driven by large correlations between factors (.70). There was a stark drop in NMI (around 0.20) from eight variables per factor to four variables per factor with these large correlations between factors. For both GLASSO and TMFG, bootEGA improved item placement in these conditions relative to empirical EGA. For all other conditions, bootEGA was comparable with empirical EGA. In sum, bootEGA provided comparable accuracy when estimating the number of dimensions and provided comparable or better item placement of items within those dimensions. TMFG particularly benefited from bootEGA.

Item Stability

To derive item stability guidelines, we collapsed across all conditions to evaluate the value at which the proportion of times an item replicating in a factor could be considered “stable” across bootstraps. To do this, we identified which simulated samples did not have NMI values equal to one (i.e., perfect item assignment) for empirical EGA. Item stability or the proportion of times an item replicated in the *simulated* dimension was then computed. We then separated items that were assigned correctly and incorrectly. Using each assignment’s item stability values, we then computed the proportion of item stability values that were above or below an incremental threshold (from 0 to 1 on .001 increments).

For the correctly assigned items, we computed the proportion of item stability values that were at or *above* the threshold. For the incorrectly assigned items, we computed the proportion of item stability values that were *below* the threshold. This approach provided inference into where items could be considered stable. Specifically, the item

315 stability value where the correctly and incorrectly assigned proportion lines crossed is the
 316 optimal threshold for maximizing items that are identified as being correctly assigned as
 317 being in the correct or incorrect population factor. The plots for each network estimation
 318 method and bootstrap type combination are presented in Figure 3.

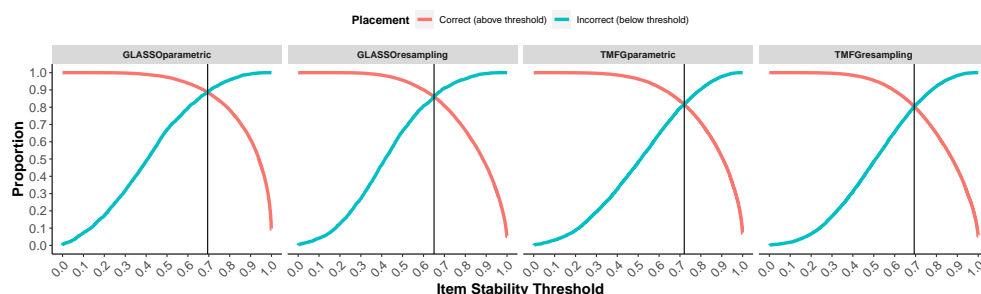


Figure 3. Item stability threshold by correct and incorrect placements of items

319 For each plot, the proportion of values that maximized the the correct and incorrect
 320 placements differed slightly for the GLASSO and TMFG methods (see vertical black lines
 321 in Figure 3). Values higher or lower than these optimal points suggest a bias towards
 322 favoring identification of correctly identifying items that are correctly placed (or correctly
 323 identifying items that are incorrectly placed). Therefore, these points represent the
 324 optimal balance in the trade-off of false positives and negatives in determining whether an
 325 item is stable or unstable. For the GLASSO method, these optimal points were at item
 326 stability values of 0.695 and 0.651 for parametric and resampling bootstraps, respectively.
 327 For the TMFG method, these optimal points were at item stability values of 0.721 and
 328 0.694 for parametric and resampling bootstraps, respectively. These results suggest that
 329 values around .65–.75 are about when variables start to become less stable.

330 Applied Example: Broad Autism Phenotype Questionnaire

331 To demonstrate how the bootEGA approach can be applied to investigate the stability of
 332 the structure of empirical data, we used data from the *Simons Simplex Collection* (SSC),
 333 a core project and resource of the Simons Foundation Autism Research Initiative that
 334 established a permanent repository of genetic and phenotypic samples from 2,600 simplex
 335 families. Each family has one child affected with an autism spectrum disorder with
 336 unaffected parents and siblings. Here, we used *Broad Autism Phenotype Questionnaire*
 337 (BAPQ) [35] data completed by mothers and fathers ($n = 5,659$) of children with an
 338 autism spectrum disorder. Before getting started with the dataset, the reader will need
 339 to install and load the latest *EGAnet* package:

```
# Install package
devtools::install_github("hfgolino/EGAnet")

# Load package
library(EGAnet)
```

340 After installing *EGAnet*, the reader should install, authenticate, download, and load data
 341 from our OSF:

```
# Install package
install.packages("osfr")

# Load package
library(osfr)

# Authenticate
?osf_auth # Instructions for how to authenticate
```

```

osf_auth("your_token_here")

# Download data
bapq_file <- osf_retrieve_file("sy9rv") %>%
  osf_download(conflicts = "skip")

# Load data
load(bapq_file$local_path)

# Rename variables with theoretical factor and item numbers
colnames(bapq) <- c(
  "A01", "P02", "R03", "P04", "A05", "R06",
  "P07", "R08", "A09", "P10", "P11", "A12",
  "R13", "P14", "R15", "A16", "P17", "A18",
  "R19", "P20", "P21", "R22", "A23", "R24",
  "A25", "R26", "A27", "A28", "P29", "R30",
  "A31", "P32", "R33", "P34", "R35", "A36"
)
row.names(bapq) <- colnames(bapq)

```

342 With the correlation matrix of BAPQ loaded, EGA can be applied. Below, we use the
 343 default EGA approach which is to use the GLASSO network estimation method with the
 344 Walktrap community detection algorithm:

```

# Estimate empirical EGA
bapq.ega <- EGA(
  bapq, # BAPQ correlation matrix
  n = 5659 # sample size
)

```

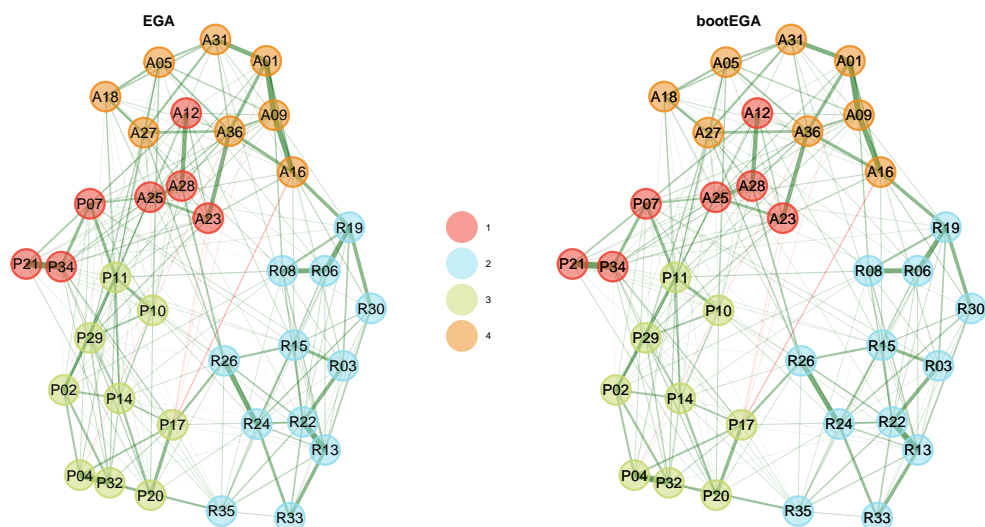


Figure 4. Dimensionality Results for EGA (left) and bootEGA (right) for the Broad Autism Phenotype Questionnaire

345 Exploratory graph analysis with the GLASSO network estimation method and the
 346 Walktrap community detection algorithm estimated four factors (Figure 4), representing
 347 the theoretical factors. Dimension 2 is consistent with the rigid dimension (R3, R6, R8,
 348 R13, R15, R19, R22, R24, R26, R30, R33, R35). Dimension 3 is consistent with the
 349 pragmatic language dimension (P2, P4, P10, P11, P14, P17, P20, P21, P29, P32, P34).
 350 Dimension 4 is consistent with the aloof dimension (A1, A5, A9, A16, A18, A27, A31,

351 A36). Dimension 1, however, represents a mix of the aloof (A12, A23, A25, A28) and
 352 pragmatic language (P7, P21, P34) dimensions.

353 Curiously, these items all reflect difficulty in social interactions with other people: A12
 354 (“People find it easy to approach me”), A23 (“I am good at making small talk”), A25
 355 (“I feel like I am really connecting with other people”), A28 (“I am warm and friendly
 356 in my interactions with others”), P07 (“I am ‘in-tune’ with the other person during
 357 conversation”), P21 (“I can tell when someone is not interested in what I am saying”) and P34
 358 (“I can tell when it is time to change topics in conversation”). While we may be
 359 interested in interpreting this dimension, it’s important to know whether these dimensions
 360 are stable and likely to reproduce in other samples. To do so, we can apply bootEGA for
 361 500 iterations:

```
# Estimate bootstrap EGA
bapq.booteqa <- bootEGA(
  bapq, # BAPQ correlation matrix
  n = 5659, # sample size
  type = "parametric", # bootstrap type
  iter = 500 # number of iterations
)
```

362 Because our dataset is a correlation matrix, the resampling (non-parametric) bootstrap
 363 option is not available. Recall that the resampling requires shuffling cases with replacement.
 364 Proceeding with the parametric bootstrap, we find that the median network structure
 365 depicted in Figure 4 has four dimensions, similar to the empirical EGA. To evaluate the
 366 stability of these dimensions, we can start first by checking the descriptive statistics:

Table 1: Descriptive statistics of BAPQ dimensions across all bootstrap replicate samples

n.Boots	median.dim	SE.dim	CI.dim	Lower.CI	Upper.CI	Lower.Quantile	Upper.Quantile
500	4	0.458	0.9	3.1	4.9	3	4

```
# Descriptive statistics
bapq.booteqa$summary.table
```

367 Here, we can see that the median is four dimensions, mirroring the empirical EGA with
 368 a relatively narrow confidence interval (CI 95% [3.10, 4.90]). To get a better idea of this
 369 distribution, we can look at the frequency of each dimension solution:

Table 2: Frequency of BAPQ dimensions across all bootstrap replicate samples

# of Factors	Frequency
3	0.298
4	0.702

```
# Frequency of dimensions
bapq.booteqa$frequency
```

370 With the frequencies, four dimensions were found 70.20% of the time or 351 of 500
 371 bootstrap replicates while three dimensions were found 29.80% of the time or 149 of 500
 372 bootstrap replicates. These results seem to suggest that the four dimension solution might
 373 be unstable. To get a better understanding of what dimensions in particular are unstable,
 374 we can compute structural consistency or how often the empirical EGA dimension is
 375 exactly replicated (identical item assignments) across the bootstrap replicates [30]:

```
# Structural consistency
```

```
bapq.dimstab <- dimensionStability(bapq.booteqa)
```

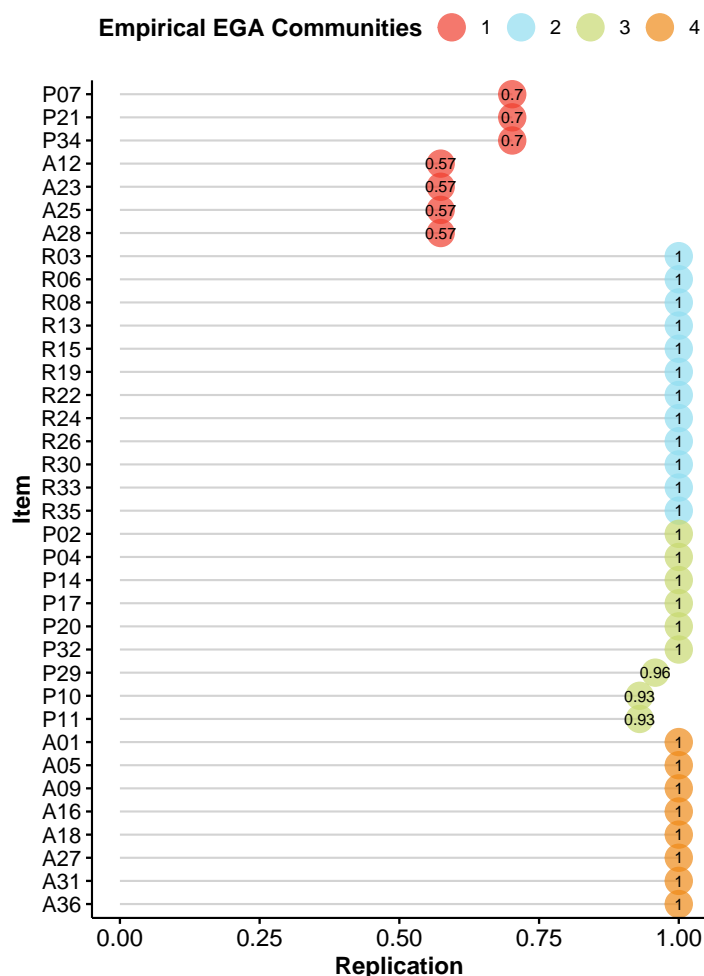


Figure 5. Item Stability of the Broad Autism Phenotype Questionnaire

```
bapq.dimstab$dimension.stability$structural.consistency
```

Table 3: Dimension stability of BAPQ items across all bootstrap replicate samples

Dimension Stability	
1	0.574
2	1.000
3	0.930
4	1.000

376 The structural consistency result shows that dimension 1, which represented the difficulty
 377 with social interactions, is very unstable. Indeed, this is confirmed when looking at the
 378 item stability values in the empirical dimensions (Figure 5). All items from dimension
 379 1 are at or below the range of .65–.75 where they can be considered unstable. Table
 380 4 shows how the item stability values across each dimension in the replicate bootstrap
 381 samples (values of zero have been removed to facilitate interpretability). To probe this
 382 instability further, we can look at how these items are replicating across all bootstrap
 383 dimensions with the following code:

```
# Item stability across all dimensions
```

```
bapq.dimstab$item.stability$item.stability$all.dimensions
```

Table 4: Item stability of BAPQ items across all bootstrap replicate samples

	1	2	3	4
P07	0.702		0.298	
P21	0.702		0.298	
P34	0.702		0.298	
A12	0.574			0.426
A23	0.574			0.426
A25	0.574			0.426
A28	0.574			0.426
R03		1		
R06		1		
R08		1		
R13		1		
R15		1		
R19		1		
R22		1		
R24		1		
R26		1		
R30		1		
R33		1		
R35		1		
P02			1	
P04			1	
P14			1	
P17			1	
P20			1	
P32			1	
P29	0.042		0.958	
P10	0.07		0.93	
P11	0.07		0.93	
A01				1
A05				1
A09				1
A16				1
A18				1
A27				1
A31				1
A36				1

Table 4 displays the item stability values across each dimension in the replicate bootstrap samples. From this table, it's clear that some items are unstable. Items A12, A23, A25, and A28 are sometimes replicating in dimension 4, which was their theoretical dimension of aloof. Items P7, P21, and P34 are sometimes replicating in dimension 3, which was theoretical dimension of pragmatic language.

These results suggest that although these items are associated with their theoretical dimension, they share enough conceptual similarity that they form a separate dimension. These unstable items are clearly causing problems with the consistency of the dimensions of the BAPQ, which are likely due to their multidimensional features—that is, sharing an underlying difficulty with social interactions. To verify this, we looked

394 at the average network loadings of these items across the bootstraps (example code:
 395 `bapq.dimstab$item.stability$item.stability$mean.loadings`; Table 5)

Table 5: Average network loadings of unstable BAPQ items across all bootstrap replicate samples

	1	2	3	4
P7	0.21	0.02	0.12	0.05
P21	0.21	0.01	0.06	0.01
P34	0.30	0.01	0.13	0.02
A12	0.18	0.02	0.02	0.09
A23	0.15	0.00	0.06	0.15
A25	0.23	0.01	0.06	0.15
A28	0.22	0.03	0.03	0.12

Note. Bold values are network loadings greater than or equal to a small effect size (0.15).

396 Based on the average network loadings, there are a couple items that have cross-loadings
 397 worth consideration (≥ 0.15) [36]: A23 (1 and 4) and A25 (1 and 4). Similarly, items P7
 398 (0.13), A28 (0.13), and P34 (0.12) have a cross-loading approaching the same threshold.
 399 These results suggest that these items are indeed multidimensional. To resolve this issue,
 400 we can remove these items and reassess the structure of the BAPQ (Figure 6 and 7).

Remove poor stability items

```
bapq.rm <- bapq[-c(7,12,21,23,25,28,34),-c(7,12,21,23,25,28,34)]
```

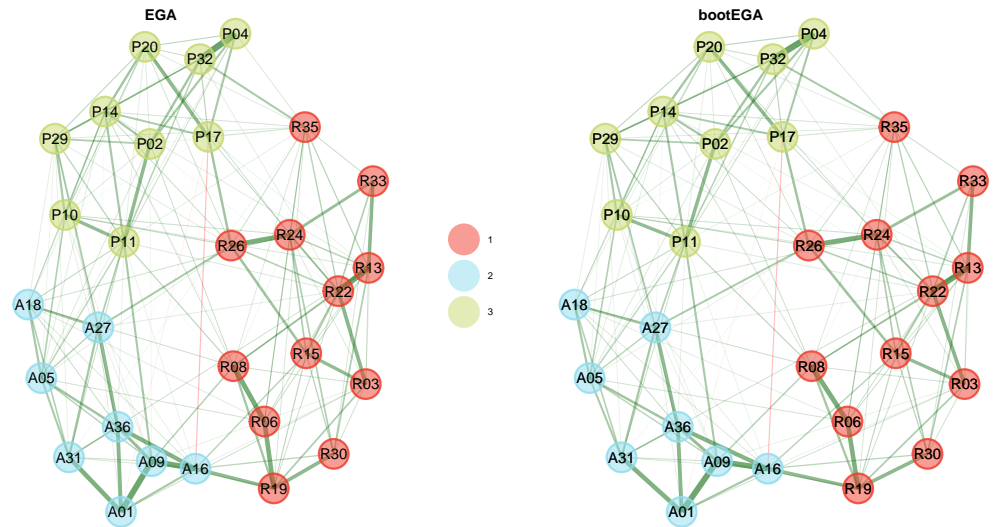


Figure 6. Dimensionality Results for EGA (left) and bootEGA (right) for the Broad Autism Phenotype Questionnaire with Unstable Items Removed

401 We can follow through with bootEGA to determine how stable the empirical EGA
 402 dimensions are after these items are removed. The bootEGA result found that three
 403 factors are being estimated in 99% of the bootstrapped samples, and Figure 7 shows
 404 that the item stability values are now nearly all 1's, demonstrating the robustness of
 405 the dimensionality solution estimated via EGA. The three factor structure estimated
 406 after removing the unstable items are very similar to the original three factor structure
 407 proposed by Hurley, Losh, Parlier, Reznick, and Piven [35]. Dimension 1 is composed
 408 of items all rigid items (R), dimension 2 of all aloof items (A), and dimension 3 of all
 409 pragmatic language items (P).

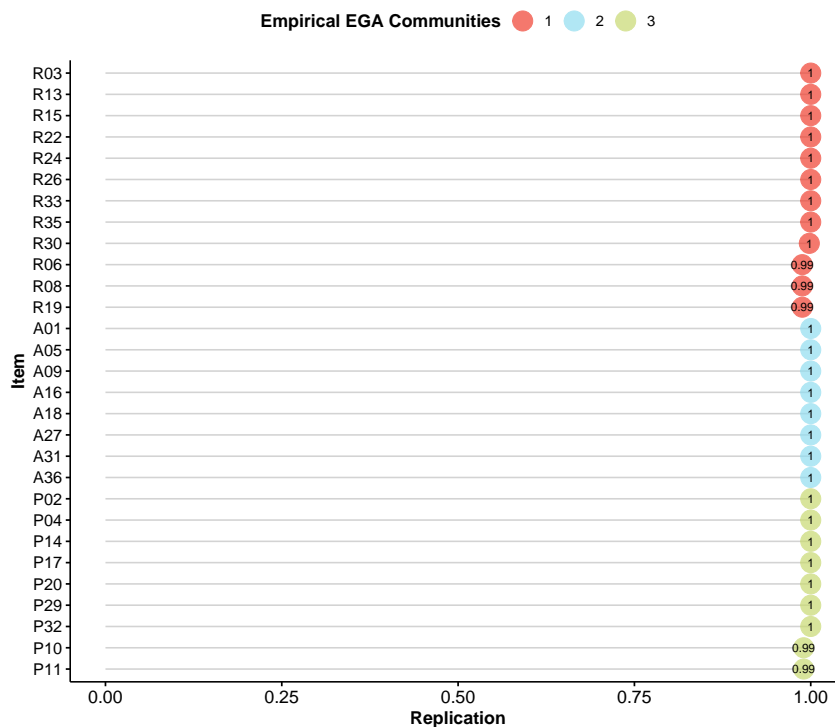


Figure 7. Item Stability of the Broad Autism Phenotype Questionnaire after removing the unstable items

This example demonstrates that multidimensional items can influence the stability of dimensions and lead to an unstable dimensional structure. Multidimensional items will replicate in two or more dimensions and can potentially be identified by examining the item stability and average network loadings across the bootstraps (e.g., Table 5). Removing these items cleans up the stability of the dimensions, leading to good structural consistency.

Discussion

In this paper, we present a novel approach for assessing the stability of dimensions in psychometric networks using EGA. Using a Monte Carlo simulation, we show that the typical structure of bootEGA is comparable to empirical EGA for estimating the number of dimensions but has as good or better item placement. From this simulation, we derived guidelines for acceptable item stability ($\geq .70$). Finally, we provide an empirical example with an associated R tutorial to demonstrate how to apply and interpret the bootEGA approach. We demonstrated that items with stability values lower than an acceptable value can lead to poor structural consistency. In our example, these items poor item stability were due to multidimensionality. We show that removing these problematic items led to improved structural consistency.

bootEGA adds another approach to the network psychometric literature for assessing the robustness of network analysis results. Previous work has applied bootstrap approaches to assess the accuracy of estimated edge weights, stability of centrality measures, and test differences between edge weights and centrality values [13]. Our approach expands reserachers' capabilities to estimate network robustness by enabling them to verify the consistency of dimensions identified by EGA. Importantly, future work will need to examine whether bootEGA provides the same benefits when data have fewer categories than we tested here (i.e., < 5). Dichotomous data, for example, may reveal greater differences in performance than the continuous and polytomous (i.e., 5-point Likert scale) data we generated.

As a part of our approach, bootEGA offers diagnostic information about the causes of poor dimensional stability. This has broad implications for scale development. For example, scales are intended to be developed to measure a single attribute. Often, attributes (measured by scales) are multifaceted with separate but related features (measured by subscales). While these features are related, there has yet to be an approach, to our knowledge, to assess the extent to which features remain cohesive [but see 37]. Such an approach is important for ensuring that each subscale is capturing a distinct feature of the attribute without being confounded by overlap with other features. In simpler language, our approach can help investigate whether the items are hanging together as researchers intend.

Our empirical example examined the BAPQ scale, which has mixed evidence for its internal structures. The BAPQ scale was intended to measure three separate but related factors of the broad autism phenotype. Previous validation studies of the BAPQ factors has shown that many items have sizable cross-loadings between factors [38]. bootEGA's dimension and item stabilities found that there were seven items belonging to the aloof and pragmatic language factors that were forming their own dimension related to difficulties with social interactions. These items were clarified using network loadings, revealing cross-loadings with considerable size. After removing these items, we identified a three-dimension structure that was structurally consistent and corroborated the theoretical factors. Future work using the BAPQ should strongly consider assessing the stability of the dimensionality of the scale before assuming the theoretical dimensions are being measured as intended.

In summary, bootEGA offers researchers several novel tools for assessing the structural integrity of their scales from the psychometric network perspective. Our approach represents both an advance in network psychometrics as well as construct validation more generally. For network analysis, researchers can assess the stability of the dimensions of the networks. For the broader construct validation literature, researchers can assess whether the structure of their subscales remain homogeneous (i.e., unidimensional) in a multidimensional context.

Materials and Methods

Data Generation

We generated data from multivariate normal factor models following the same approach as Golino et al. [2]. First, the reproduced population correlation matrix was computed:

$$\mathbf{R}_R = \mathbf{\Omega}\mathbf{\Phi}\mathbf{\Omega}',$$

where \mathbf{R}_R is the reproduced population correlation matrix, $\mathbf{\Omega}$ is the k (variables) \times r (factors) factor loading matrix,¹ 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 $\mathbf{\Omega}$ was re-generated and the same procedure was followed until these criteria are met. Finally, the sample data matrix of continuous variables was computed:

¹ The traditional notation for factor loadings matrix is $\mathbf{\Lambda}$; however, to avoid confusion with the GLASSO's lambda parameter (λ), we use omega ($\mathbf{\Omega}$).

$$\mathbf{X} = \mathbf{Z}\mathbf{U},$$

where \mathbf{Z} is a matrix of random multivariate normal data with rows equal to the sample size and columns equal to the number of variables.

To generate polytomous data, each continuous variable was categorized into five categories, resembling a 5-point Likert scale, with a random skew ranging from -2 to 2 on a 0.5 interval from a random uniform distribution following the approach of Garrido, Abad, and Ponsoda [39,40].

Design

Our goal for this simulation was twofold. First, we wanted to evaluate whether the typical structure of bootEGA improved the accuracy of the number of dimensions and placement of items over empirical EGA. Second, we wanted to define the item stability guidelines. For this latter goal, we used the data from the largest sample size condition (1000) only to get the most robust guidelines.

The data were simulated from a multidimensional multivariate normal distribution with factor loadings for each item randomly drawn from a uniform distribution ranging from .40-.70. Cross-loadings were also generated following a random normal distribution with a mean of zero and a standard deviation of .10. The data categories (continuous and polytomous), variables per factor (4 and 8), number of factors (2 and 4), sample size (250, 500, and 1000), and correlations between factors (.00, .30, .50, and .70) were manipulated. In total, a $2 \times 2 \times 2 \times 3 \times 4$ (data categories, variables per factor, number of factors, sample size, factor correlations) between-subject design was implemented, resulting in 96 conditions. For each condition, 100 datasets were generated, resulting in 9600 datasets. We examined both EGA network estimation methods (GLASSO and TMFG) as well as both bootstrap methods (parametric and resampling). The *bootEGA* technique was implemented using 500 bootstraps for each condition. In total, 19,200,000 datasets were generated (including the bootstrap replicates).

Network Estimation Methods

GLASSO

The GLASSO uses the LASSO, which is a statistical regularization technique that reduces parameter estimates, with some estimates becoming exactly zero [for the mathematical notation, see 41]. The aim of this technique is to achieve a sparse network model—non-relevant edges are removed from the model, leaving only a subset of relevant (not necessarily significant) edges.

This sparsity is controlled by a parameter called *lambda* (λ). Lower values of lambda remove fewer edges, increasing the possibility of including spurious associations, while larger values of lambda remove more edges, increasing the possibility of removing relevant edges. When $\lambda = 0$, then the estimates are equal to the ordinary least squares solution (i.e., the partial correlation matrix). This parameter is thus an important part of model selection, striking a balance between sensitivity (i.e., selecting relevant edges that are truly relevant) and specificity (i.e., removing edges that are truly not relevant).

The popular approach in the network psychometrics literature is to compute models across several values of λ (usually 100) and to select the model that minimizes the EBIC. The EBIC model selection uses a hyperparameter (γ) to control how much it prefers simpler models (i.e., models with fewer edges) [14]. 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. In the psychometric network literature, this approach has been termed EBICglasso and is applied via the *qgraph* package [42] in R. For continuous

524 data, Pearson’s correlations were computed; for polytomous data, polychoric correlations
525 were computed.

526 Following the EGA approach [2], the minimum λ value is set to .01, which is slightly
527 larger than the default of .001. This larger value is selected to reduce the prevalence of
528 false positive edges in the network. Next, the γ value is set to .50, which is the default;
529 however, it is iteratively decreased by .25, until reaching zero, based on whether any
530 one node in the network is disconnected. If γ reaches zero, then the network is used
531 regardless of whether any nodes are disconnected.

532 TMFG

533 This study applied the TMFG [15,43], which applies a structural constraint on the
534 zero-order correlation matrix. This constraint restrains the network to retain a certain
535 number of edges ($3n - 6$, where n is the number of nodes). This network is comprised
536 of three- and four-node cliques (i.e., sets of connected nodes; a triangle and tetrahedron,
537 respectively).

538 Network estimation starts with a tetrahedron that is comprised of the four nodes that
539 have the high sum of correlations that are greater than the average correlation in the
540 correlation matrix. Next, the algorithm identifies a node that is not connected in the
541 network and maximizes its sum of correlations to three nodes already in the network.
542 This node is then connected to those three nodes. This process continues iteratively until
543 every node is connected in the network.

544 The resulting network is a planar network or a network that could be drawn such that
545 no edges are crossing [44]. One property of these networks is that they form a “nested
546 hierarchy” such that its constituent elements (3-node cliques) form sub-networks in the
547 overall network [17].

548 Walktrap Algorithm

549 To define the random walk, let \mathbf{A} be a square matrix of edge weights (e.g., partial
550 correlations) in the network, where \mathbf{A}_{ij} is the strength of the (partial) correlation between
551 node i and j and a node’s strength is the sum of node i ’s connections to its neighbors
552 $NS = \sum_j \mathbf{A}_{ij}$. The steps move from one node to another randomly and uniformly using
553 a transition probability, $\mathbf{P}_{ij} = \frac{\mathbf{A}_{ij}}{NS(i)}$, which forms the transition matrix, \mathbf{P} .

554 To determine the communities that the nodes belong to, the transition matrix is used to
555 compute a distance metric, r , which measures the structural similarity between nodes.

556 This structural similarity is defined as [20]:

$$r_{ij} = \sqrt{\sum_{k=1}^n \frac{(\mathbf{P}_{ik} - \mathbf{P}_{jk})^2}{NS(k)}}$$

557 This distance can be generalized to the distance between nodes and communities by
558 beginning the random walk at a random node in a community, C . This can be defined as:

$$\mathbf{P}_{Cj} = \frac{1}{|C|} \sum_{i \in C} \mathbf{P}_{ij}.$$

559 Finally, this can be further generalized to the distance between two communities:

$$r_{C_1 C_2} = \sqrt{\sum_{k=1}^n \frac{(\mathbf{P}_{C_1 k} - \mathbf{P}_{C_2 k})^2}{NS(k)}}$$

where this definition is consistent with the distance between nodes in the network.

The algorithm begins by having each node as a cluster (i.e., n clusters). The distances, r , are computed between all adjacent nodes, and the algorithm then begins to iteratively choose two clusters. These two clusters chosen are then merged into a new cluster, updating the distances between the node(s) and cluster(s) with each merge (in each $k - n - 1$ steps).

Clusters are only merged if they are adjacent to one another (i.e., an edge between them). The merging method is based on Ward's agglomerative clustering approach [45] that depends on the estimation of the squared distances between each node and its community (σ_k), for each k steps of the algorithm. Since computing σ_k is computationally expensive, Pons and Latapy [20] adopted an efficient approximation that only depends on the nodes and the communities rather than the k steps. The approximation seeks to minimize the variation of σ that would be induced if two clusters (C_1 and C_2) are merged into a new cluster (C_3):

$$\Delta\sigma(C_1, C_2) = \frac{1}{n} \left(\sum_{i \in C_3} r_{iC_3}^2 - \sum_{i \in C_1} r_{iC_1}^2 - \sum_{i \in C_2} r_{iC_2}^2 \right).$$

Because Ward's approximation adopted by Pons and Latapy [20] only merges adjacent clusters, the total number of times $\Delta\sigma$ is updated is not very large, and the resulting values can be stored in a balanced tree. A sequence of P_k partitions into clusters ($1 \leq k \leq n$, being n the total number of nodes) is obtained. The best number of clusters is defined as the partition that maximizes modularity.

Modularity is a measure that was proposed by Newman [21] to identify meaningful clusters in networks and is calculated as follows. Let j and k be two clusters in a network with m and n nodes. If the number of edges between clusters is p , then one-half of fraction of the edges linking j and k is $e_{jk} = \frac{1}{2}p$, so that the total fraction of edges between the two clusters is $e_{jk} + e_{kj}$ [21]. On the other hand, e_{jj} represents the fraction of edges that fall within cluster j , whose sum equals one: $\sum_j e_{jj} = 1$. Newman [21] points out that a division of networks into clusters is meaningful if the value of the sums of e_{jj} and e_{kk} is maximized.

Statistical Analyses

To evaluate the performance of the *bootEGA* algorithm, we used two complementary criteria: the accuracy or percentage of correct estimates (PC) and the normalized mutual information (NMI). The accuracy criterion varies between zero (0% correct estimates) and one (100% correct estimates). PC is defined as:

$$PC = \frac{\sum C}{N}, \text{ for } C = \begin{cases} 1 & \text{if } \hat{\theta} = \theta \\ 0 & \text{if } \hat{\theta} \neq \theta \end{cases},$$

where $\hat{\theta}$ is the estimated number of factors, θ is the population number of factors, and N is the number of sample data matrices simulated.

The normalized mutual information measures how well the estimated assignment of items per factor reflects the simulated structure. A value of one indicates that all items are assigned to the correct factor in the population, and a value of zero indicates that all items are assigned to incorrect factors in the population. NMI defines a confusion matrix, N , where the rows correspond to the population dimensions and the columns correspond to the estimated dimensions. The element, C_{ij} , refers to the number of items that are found in population factor i that are in the estimated factor j . Using the information-theoretic measure of mutual information, this defines NMI as:

$$NMI = \frac{-2 \sum_{i=1}^{C_A} \sum_{j=1}^{C_B} N_{ij} \log(N_{ij}N / N_i.N_j)}{\sum_{i=1}^{C_A} N_i. \log(N_i./N) + \sum_{j=1}^{C_B} N.j \log(N.j/N)},$$

where C_A is the number of population dimensions and C_B is the number of estimated dimensions.

We compared the empirical EGA results with the median network structure of the bootEGA replicate samples on their PC and NMI. We applied ANOVAs for both PC and NMI with partial eta squared (η_p^2) effect sizes following Cohen's [46] guidelines: .01 (small), .06 (moderate), and .14 (large).

To determine an appropriate cut-off for the item stability statistic of how often an item replicates within a dimension is indicative of "good stability." To do this, we used the bootEGA results from Simulation 1 and computed the item stability statistics (specifically, replication proportions across dimensions). Then, using the empirical EGA dimensionality solutions that did not have proper item assignment (i.e., $NMI < 1$), we identified the stability of the items that were correctly and incorrectly assigned to the *population* factor structure. This approach allowed us to determine the stability of items that are appropriately identified in the correct dimensional solution versus when they are not. We then plotted the proportion of values across an incrementally increasing threshold (.001).

Broad Autism Phenotype Questionnaire

For our second applied example, we used data completed as a part of the Simons Foundation Autism Research Initiative's Simplex Collection (<https://www.sfari.org/>). This data included the BAPQ [35], which was completed by 5,659 individuals (fathers and mothers of a child affected with an autism spectrum disorder). The original dimensionality structure proposed by Hurley, Losh, Parlier, Reznick, and Piven [35] has three factors which capture different aspects of the broad autism phenotype: aloof personality (represents a limited interest in or enjoyment of social interactions), rigid personality (refers to resistance, and/or difficulty adapting, to change) and pragmatic language (deficits in the social use of language leading to difficulties with effective communication and/or conversational reciprocity). Each factor has 12 items (36 items in total).

Data Analyses

We used R (version 4.0.5) [47] for all of our analyses. The functions for all analyses used in this paper are housed in the *EGAnet* package (version 0.9.9) [48]. *EGAnet* uses the packages *ggplot2* (version 3.3.3) [49] and *ggpubr* (version 0.4.0) [50] for plots and network visualization. Network estimation methods are implemented via the packages *qgraph* (GLASSO; version 1.6.9) [42] and *NetworkToolbox* (TMFG; version 1.4.2) [51]. Finally, the Walktrap algorithm is implemented using the *igraph* package (version 1.2.6) [52]. All data, code, and *Rmarkdown* files used in the present paper can be found on the Open Science Framework: <https://osf.io/wxdk7/>.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, A.P.C and H.G.; methodology, A.P.C. and H.G.; software, A.P.C. and H.G.; validation, A.P.C. and H.G.; formal analysis, A.P.C. and H.G.; investigation, A.P.C. and H.G.; resources, H.G.; data curation, A.P.C.; writing—original draft preparation, A.P.C.; writing—review and editing, A.P.C. and H.G.; visualization, A.P.C. and H.G.; supervision, H.G.; funding acquisition, H.G. All authors have read and agreed to the published version of the manuscript.", please turn to the [CRediT taxonomy](#) for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: This project was partially funded by the University of Virginia Support Transformative Autism Research initiative.

648 **Institutional Review Board Statement:** Not applicable

649 **Informed Consent Statement:** Not applicable

650 **Data Availability Statement:** All data, code, and *Rmarkdown* files used in the present paper
651 can be found on the Open Science Framework: <https://osf.io/wxdk7/>.

652 **Acknowledgments:** We would like to thank the University of Virginia Support Transformative
653 Autism Research initiative for their funding support and the Simons Foundation Autism Research
654 Initiative for providing us access to their data.

655 **Conflicts of Interest:** The authors declare no conflict of interest.

656 Abbreviations

657 The following abbreviations are used in this manuscript:

658

EGA	Exploratory Graph Analysis
bootEGA	Bootstrap Exploratory Graph Analysis
GLASSO	Graphical Least Absolute Shrinkage and Selection Operator
LASSO	Least Absolute Shrinkage and Selection Operator
GGM	Gaussian Graphical Model
EBIC	Extended Bayesian Information Criterion
TMFG	Triangulated Maximally Filtered Graph
659 PA	Parallel Analysis
PAF	Principal Axis Factoring
PCA	Principal Component Analysis
ARI	Adjusted Rand Index
PC	Percent Correct
NMI	Normalized Mutual Information
SSC	Simons Simplex Collection
BAPQ	Broad Autism Phenotype Questionnaire

References

1. Golino, H.; Epskamp, S. Exploratory Graph Analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS ONE* **2017**, *12*, e0174035. doi:10.1371/journal.pone.0174035.
2. Golino, H.; Shi, D.; Christensen, A.P.; Garrido, L.E.; Nieto, M.D.; Sadana, R.; Thiagarajan, J.A.; Martinez-Molina, A. Investigating the performance of Exploratory Graph Analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods* **2020**, *25*, 292–320. doi:10.1037/met0000255.
3. Fortunato, S. Community Detection in Graphs. *Physics Reports* **2010**, *486*, 75–174. doi:10.1016/j.physrep.2009.11.002.
4. Borsboom, D.; Robinaugh, D.J.; Group, P.; Rhemtulla, M.; Cramer, A.O.J. Robustness and replicability of psychopathology networks. *World Psychiatry* **2018**, *17*, 143–144. doi:10.1002/wps.20515.
5. Forbes, M.K.; Wright, A.G.C.; Markon, K.E.; Krueger, R.F. Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology* **2017**, *126*, 969–988. doi:10.1037/abn0000276.
6. Forbes, M.K.; Wright, A.G.; Markon, K.E.; Krueger, R.F. Quantifying the reliability and replicability of psychopathology network characteristics. *Multivariate Behavioral Research* **2019**. doi:10.1080/00273171.2019.1616526.
7. Fried, E.I.; Cramer, A.O.J. Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science* **2017**, *12*, 999–1020. doi:10.1177/1745691617705892.
8. Friedman, J.; Hastie, T.; Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **2008**, *9*, 432–441. doi:10.1093/biostatistics/kxm045.
9. Friedman, J.; Hastie, T.; Tibshirani, R. *glasso: Graphical lasso – Estimation of Gaussian graphical models*, 2014. R package version 1.8.
10. Lauritzen, S.L. *Graphical Models*; Clarendon Press: Oxford, UK, 1996.
11. Tibshirani, R. Regression Shrinkage and Selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **1996**, pp. 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.
12. Chen, J.; Chen, Z. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **2008**, *95*, 759–771. doi:10.1093/biomet/asn034.

13. Epskamp, S.; Borsboom, D.; Fried, E.I. Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods* **2018**, *50*, 195–212. doi:10.3758/s13428-017-0862-1.
14. Foygel, R.; Drton, M. Extended Bayesian information criteria for Gaussian graphical models. Advances in neural information processing systems; Lafferty, J.D.; Williams, C.K.I.; Shawe-Taylor, J.; Zemel, R.S.; Culotta, A., Eds., 2010, pp. 604–612.
15. Massara, G.P.; Di Matteo, T.; Aste, T. Network filtering for big data: Triangulated Maximally Filtered Graph. *Journal of Complex Networks* **2016**, *5*, 161–178. doi:10.1093/comnet/cnw015.
16. Christensen, A.P.; Cotter, K.N.; Silvia, P.J. Reopening openness to experience: A network analysis of four openness to experience inventories. *Journal of Personality Assessment* **2019**, *101*, 574–588. doi:10.1080/00223891.2018.1467428.
17. Song, W.M.; Di Matteo, T.; Aste, T. Hierarchical Information Clustering by Means of Topologically Embedded Graphs. *PLoS One* **2012**, *7*, e31929. doi:10.1371/journal.pone.0031929.
18. Christensen, A.P.; Garrido, L.E.; Golino, H. Comparing community detection algorithms in psychological data: A Monte Carlo simulation. *PsyArXiv* **2021**. doi:10.31234/osf.io/hz89e.
19. Gates, K.M.; Henry, T.; Steinley, D.; Fair, D.A. A Monte Carlo evaluation of weighted community detection algorithms. *Frontiers in Neuroinformatics* **2016**, *10*, 45. doi:10.3389/fninf.2016.00045.
20. Pons, P.; Latapy, M. Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications* **2006**, *10*, 191–218. doi:10.7155/jgaa.00185.
21. Newman, M.E.J. Modularity and community structure in networks. *Proceedings of the national academy of sciences* **2006**, *103*, 8577–8582. doi:10.1073/pnas.0601602103.
22. Guttman, L. Some necessary conditions for common-factor analysis. *Psychometrika* **1954**, *19*, 149–161. doi:10.1007/BF02289162.
23. Kaiser, H.F. The application of electronic computers to factor analysis. *Educational and psychological measurement* **1960**, *20*, 141–151. doi:10.1177/001316446002000116.
24. Velicer, W.F. Determining the number of components from the matrix of partial correlations. *Psychometrika* **1976**, *41*, 321–327. doi:10.1007/BF02293557.
25. Horn, J.L. A rationale and test for the number of factors in factor analysis. *Psychometrika* **1965**, *30*, 179–185. doi:10.1007/BF02289447.
26. Blondel, V.D.; Guillaume, J.L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, *2008*, P10008. doi:10.1088/1742-5468/2008/10/P10008.
27. Clauset, A.; Newman, M.E.J.; Moore, C. Finding community structure in very large networks. *Physical review E* **2004**, *70*, 066111. doi:10.1103/PhysRevE.70.066111.
28. Efron, B.; Tibshirani, R.J. *An introduction to the bootstrap*; CRC press: Boca Raton, FL, 1994.
29. Jung, K.; Lee, J.; Gupta, V.; Cho, G. Comparison of bootstrap confidence interval methods for GSCA using a Monte Carlo simulation. *Frontiers in Psychology* **2019**, *10*, 2215. doi:10.3389/fpsyg.2019.02215.
30. Christensen, A.P.; Golino, H.; Silvia, P.J. A psychometric network perspective on the validity and validation of personality trait questionnaires. *European Journal of Personality* **2020**, *34*, 1095–1108. doi:10.1002/per.2265.
31. Hubert, L.; Arabie, P. Comparing partitions. *Journal of classification* **1985**, *2*, 193–218. doi:10.1007/BF01908075.
32. Christensen, A.P.; Garrido, L.E.; Golino, H. Unique variable analysis: A novel approach for detecting redundant variables in multivariate data. *PsyArXiv* **2020**. doi:10.31234/osf.io/4kra2.
33. Montoya, A.K.; Edwards, M.C. The poor fit of model fit for selecting number of factors in exploratory factor analysis for scale evaluation. *Educational and Psychological Measurement* **2020**. doi:10.1177/0013164420942899.
34. Auerswald, M.; Moshagen, M. How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods* **2019**, *24*, 468–491. doi:10.1037/met0000200.
35. Hurley, R.S.E.; Losh, M.; Parlier, M.; Reznick, J.S.; Piven, J. The broad autism phenotype questionnaire. *Journal of Autism and Developmental Disorders* **2007**, *37*, 1679–1690. doi:10.1007/s10803-006-0299-3.
36. Christensen, A.P.; Golino, H. On the equivalency of factor and network loadings. *Behavior Research Methods* **2021**, *53*, 1563–1580. doi:10.3758/s13428-020-01500-6.
37. Waller, N.G.; DeYoung, C.G.; Bouchard, T.J. The recaptured scale technique: A method for testing the structural robustness of personality scales. *Multivariate Behavioral Research* **2016**, *51*, 433–445. doi:10.1080/00273171.2016.1157753.
38. Sasson, N.J.; Lam, K.S.; Childress, D.; Parlier, M.; Daniels, J.L.; Piven, J. The Broad Autism Phenotype Questionnaire: Prevalence and diagnostic classification. *Autism Research* **2013**, *6*, 134–143. doi:10.1002/aur.1272.
39. Garrido, L.E.; Abad, F.J.; Ponsoda, V. Performance of Velicer’s minimum average partial factor retention method with categorical variables. *Educational and Psychological Measurement* **2011**, *71*, 551–570. doi:10.1177/0013164410389489.
40. Garrido, L.E.; Abad, F.J.; Ponsoda, V. A new look at Horn’s parallel analysis with ordinal variables. *Psychological Methods* **2013**, *18*, 454–474. doi:10.1037/a0030005.
41. Epskamp, S.; Fried, E.I. A tutorial on regularized partial correlation networks. *Psychological Methods* **2018**, *23*, 617–634. doi:10.1037/met0000167.
42. Epskamp, S.; Cramer, A.O.J.; Waldorp, L.J.; Schmittmann, V.D.; Borsboom, D. qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software* **2012**, *48*, 1–18. doi:10.18637/jss.v048.i04.

-
43. Christensen, A.P.; Kenett, Y.N.; Aste, T.; Silvia, P.J.; Kwapil, T.R. Network structure of the Wisconsin Schizotypy Scales–Short Forms: Examining psychometric network filtering approaches. *Behavior Research Methods* **2018**, *50*, 2531–2550. doi:10.3758/s13428-018-1032-9.
 44. Tumminello, M.; Aste, T.; Di Matteo, T.; Mantegna, R.N. A Tool for Filtering Information in Complex Systems. *Proceedings of the National Academy of Sciences* **2005**, *102*, 10421–10426. doi:10.1073/pnas.0500298102.
 45. Ward, J.H. Hierarchical clustering to optimise an objective function. *Journal of the American Statistical Association* **1963**, *58*, 238–244.
 46. Cohen, J. *Statistical power analysis for the behavioural sciences*, 2 ed.; Routledge: New York, NY, 1988. doi:10.4324/9780203771587.
 47. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
 48. Golino, H.; Christensen, A.P. *EGAnet: Exploratory Graph Analysis– A Framework for Estimating the Number of Dimensions in Multivariate Data Using Network Psychometrics*, 2020. R package version 0.9.9.
 49. Wickham, H. *ggplot2: Elegant graphics for data analysis*; Springer: New York, NY, 2016.
 50. Kassambara, A. *ggpubr: ‘ggplot2’ Based Publication Ready Plots*, 2018. R package version 0.2.
 51. Christensen, A.P. NetworkToolbox: Methods and Measures for Brain, Cognitive, and Psychometric Network Analysis in R. *The R Journal* **2018**, *10*, 422–439. doi:10.32614/RJ-2018-065.
 52. Csardi, G.; Nepusz, T. The igraph Software Package for Complex Network Research. *InterJournal, Complex Systems* **2006**, *1695*, 1–9.