MULTIVARIATE BEHAVIORAL RESEARCH
2023, VOL. 58, NO. 2, 311–339

Bayesian Uncertainty Estimation for Gaussian Graphical Models and Centrality Indices

Joran Jongerling[a], Sacha Epskamp[b,c], and Donald R. Williams[d]

[a]Department of Methodology and Statistics, Tilburg School of Social and Behavioral Sciences, Tilburg University; [b]Department of Psychology, Faculty of Social and Behavioral Sciences, University of Amsterdam; [c]Centre for Urban Mental Health, University of Amsterdam; [d]Department of Psychology, University of California

ABSTRACT

In the network approach to psychopathology, psychological constructs are conceptualized as networks of interacting components (e.g., the symptoms of a disorder). In this network view, interest is on the degree to which symptoms influence each other, both directly and indirectly. These direct and indirect influences are often captured with centrality indices, however, the estimation method often used with these networks, the frequentist graphical LASSO (GLASSO), has difficulty estimating (uncertainty in) these measures. Bayesian estimation might provide a solution, as it is better suited to deal with bias in the sampling distribution of centrality indices. This study therefore compares estimation of symptom networks with Bayesian GLASSO- and Horseshoe priors to estimation using the frequentist GLASSO using extensive simulations. Results showed that the Bayesian GLASSO performed better than the Horseshoe, and that the Bayesian GLASSO outperformed the frequentist GLASSO with respect to bias in edge weights, centrality measures, correlation between estimated and true partial correlations, and specificity. Sensitivity was better for the frequentist GLASSO, but performance of the Bayesian GLASSO is usually close. With respect to uncertainty in the centrality measures, the Bayesian GLASSO shows good coverage for strength and closeness centrality, but uncertainty in betweenness centrality is estimated less well.

KEYWORDS

Network; LASSO; horseshoe; centrality; Bayesian

Introduction

In recent years, psychological research is increasingly adopting a network perspective to psychological behavior (Borsboom, 2008; Borsboom et al., 2011; Borsboom & Cramer, 2013; Cramer et al., 2010; Schmittmann et al., 2013) in which psychological constructs are conceptualized as networks of interacting components referred to as nodes in network literature. Usually in psychological networks, these nodes represent observed variables (e.g., symptoms), while connections between nodes (edges) represent statistical relationships between the behaviors (Epskamp et al., 2017). Psychological networks are often estimated using Gaussian Graphical Models (GGM; Costantini et al., 2015; Lauritzen, 1996), which assume that the data are multivariate normally distributed, and in which edges can be interpreted as partial correlation coefficients, that is, correlations between two nodes after conditioning on all other nodes in the network, which are proportional to regression coefficients (Epskamp et al., 2017; Epskamp & Fried, 2018). Currently, a popular form of regularized GGM estimation is by using the graphical LASSO (GLASSO) (Epskamp & Fried, 2018; Friedman et al., 2008), which limits the sum of the absolute edge weights, and therefore shrinks small estimates toward zero. As such, the GLASSO returns a sparse network model, in which only a relatively small number of edges are used to explain the covariation structure in the data (Epskamp et al., 2017). Because of this sparsity, the estimated models become more interpretable. Regularized estimation is not (always) necessary (especially with larger sample sizes), and

unregularized estimation of GGMs is also possible (Liang et al., 2015; Williams et al., 2019). Regularized estimation is often used however, and is the estimation method on which we will focus in this study.

There are currently no real significance tests for edge weights in regularized networks (although they do exist for unregularized networks, for example in the *psychonetrics* R-package (Epskamp, 2020). Instead, since regularization sets small edges equal to zero, the presence of an edge alone is taken as evidence that the two nodes are related to each (after conditioning on all other nodes in the network). The "importance" of nodes is often subsequently operationalized as how connected a node is to all others, either directly (*strength centrality*: the sum of the absolute values of all edges of a node (i.e., the partial correlations between the node and all other nodes)) or indirectly (*closeness centrality*: the inverse of the sum of largest indirect effects between nodes; *betweenness centrality*: the number of shortest paths between two other nodes that a node is part of)—although whether connectedness is indeed a good measure of importance, and whether it is always informative is under debate (see for example, Bringmann et al. (2019)). Yet, proper inference based on network models requires taking the accuracy of the estimates of edges and centrality measures into account, both to get an idea about the range of plausible value for these parameters, and to be able to determine which nodes can be considered as different from each other. Epskamp et al. (2017) introduced a method, included in the *bootnet* R-package (Epskamp et al., 2017), to do this based on bootstrapping. However, while their method allows the estimation of intervals representing likely values of edges, called *bootstrapped confidence intervals*, these intervals are not the same as regular confidence intervals and can't be used to test the null hypothesis of no relation when regularization is used. In addition, simulations showed that it is difficult to get unbiased estimates and 95% confidence intervals for centrality indices. This is due to the instability in centrality indices caused by sampling variation and due to bias in their sampling distributions (i.e., strength centrality is calculated using the absolute value of edge weights which leads to skewed distributions, also see the supplementary material of Epskamp et al. (2017)). In addition, their results showed that constructing bootstrapped CIs on very low significance levels is not feasible with a limited number of bootstrap samples (Epskamp et al., 2017).

In this paper, we therefore investigate Bayesian estimation of GGMs using a Bayesian version of the GLASSO (Wang, 2012) and using a Graphical Horseshoe prior (Li et al., 2019) for the partial correlation matrix. We chose these two Bayesian estimation methods because they are popular in the social science literature and therefore often used. In addition, we wanted to test alternative methods against the regularized estimation method that is quite common in the social sciences, which is the frequentist GLASSO. A Bayesian GLASSO is the most natural comparison method for this often used frequentist GLASSO in our opinion. Since the Bayesian estimation of GGMs provides posterior distributions for all sampled parameters and allows easy estimation of transformations and/or functions of these parameters, these methods could lead to measures of centrality that are less biased. For example, as mentioned above, strength centrality is calculated using the sum of the *absolute* values of the edge weights of a node (so positive and negative nodes don't cancel out), but this distorts the distribution of edges, and therefore the distribution of sums of (absolute values) of these edges as well (strength centrality) as well (see the left panel of Figure 1).

With Bayesian analysis it's easy to use an alternative for absolute values. We could, for example, simply shift the posterior distributions of edge weights to be centered around positive values. If a posterior is centered around −.20 (the left distribution in both panels of Figure 1), we can shift the entire distribution .4 points to the right to center it around a *positive* value of .20 (right distribution in the right panel of Figure 1). As can be seen in Figure 1, shifting the entire posterior prevents distortion of the distributions of an edge weight (the right distribution in the right panel of Figure 1 is the exact same shape as the left distribution in that panel), unlike taking absolute values (the two distributions in the left panel of Figure 1 are clearly different). If we apply an appropriate shift to each individual edge (to make sure they are all centered around the positive 'version' of their individual point estimates), we can still prevent positive and negative (point estimates of) edges from canceling out when calculating strength centrality, but do so while distorting the distributions of the edges less. This could make the sum of these shifted posterior distributions a better behaved estimate of strength centrality than one based on absolute values. An extensive simulation study will therefore be undertaken to determine the ability of our Bayesian estimation methods to accurately estimate centrality measures and their uncertainty. In addition, to make sure that being able to estimate centralities and there uncertainty does not come at the cost of precision in estimation of individual edges, we also compare the Bayesian GLASSO and Horseshoe against the "standard" frequentist GLASSO (with tuning parameter selection based on the Extended Bayesian Information Criterion (EBIC) as implemented in the R-package

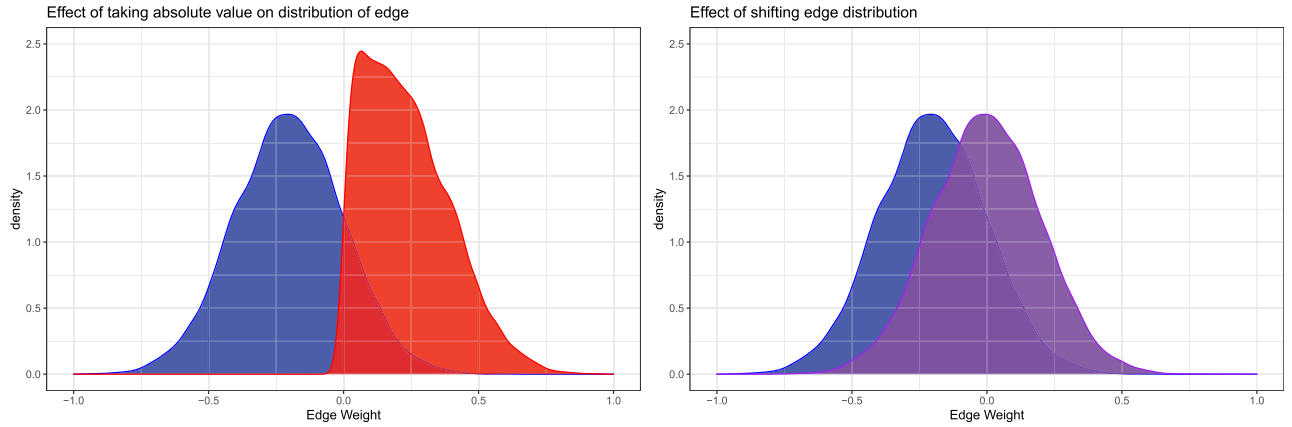


Figure 1. Effect of taking absolute values vs. shifting posteriors on the distribution of edges. (a) Psych Network: Network based on the bfi-data. (b) RED Network: Random Network with density equal to the Psych Network. (c) RHD Network: Random Network with density half that of the Psych Network.

bootnet (Epskamp et al., 2017) with respect to bias in edges, sensitivity, and specificity. This article is structured as follows. In the next section we will fist discuss the Bayesian GLASSO and Horseshoe estimation methods for GGMs. Then we will describe the different methods for estimating strength, closeness, and betweenness centrality that we will be testing. In the third section we will describe our simulation study, the results of which are presented in section four. We end with a discussion.

The Bayesian GLASSO and Horseshoe, and the estimation methods for the different centrality measures are incorporated into the R-package *BUEG* (Bayesian Uncertainty Estimation for GGMs) which is available on OSF (https://osf.io/9kjxv/).

Bayesian estimation of GGMs

Assume we have observations from N individuals on P multivariate normally distributed variables $\mathbf{y_p}(p = 1, ..., P)$. If the means of these variables are equal to 0, the distribution of these data are given by,

$$\mathbf{y} \sim \text{MVN}(\mathbf{0}, \mathbf{\Sigma}), \tag{1}$$

where $\mathbf{y}$ is a N by P matrix containing all item responses of the N individual on the P variables $\mathbf{y_p}$ (i.e., the vectors $\mathbf{y_p}$ make up the columns of matrix $\mathbf{y}$), and $\mathbf{\Sigma}$ is the variance-covariance matrix. The inverse of the variance-covariance matrix, $\mathbf{K}$, is called the precision matrix, which, after standardizing and inverting the signs of its elements, contains the partial correlations between the P random variables on its off-diagonal elements. These partial correlations can be displayed as a weighted network, in which each variable $\mathbf{y_p}$ represents a node, and the partial correlations between variables show up as connections (edges) between the respective nodes. If a partial correlation between two variables is equal to 0, these two variables are conditionally independent given all other variables, and the nodes of these variables are unconnected in the graph of the weighted network. A challenge in precision matrix estimation is that the number of free parameters grows quadratically with the number of variables, which is why the precision matrix is often assumed to be sparse (i.e., some elements in $\mathbf{K}$ are expected to be zero even though every element of $\mathbf{\Sigma}$ may be non-zero).

Estimation of such a sparse model, termed a Gaussian Graphical Model (GGM; Epskamp et al., 2017) when data are multivariate normally distributed, can be achieved by penalizing the likelihood. A popular penalized estimation method is the Graphical LASSO (GLASSO) proposed by Friedman et al. (2008), which uses the LASSO penalization (Tibshirani, 1996) and can be written as,

$$\log\left(\det\mathbf{K}\right) - \text{tr}(\mathbf{SK}) - \sum_{j,k} \phi_\lambda(|\kappa_{jk}|), \tag{2}$$

where $\mathbf{S}$ is the covariance matrix, κ_{jk} is the entry on the jth row and kth column of $\mathbf{K}$ (proportional to the partial correlation between variables j and k), $\phi_\lambda(|\kappa_{jk}|) = \lambda|\kappa_{jk}|$ is the ℓ_1 penalty, and λ (with $\lambda > 0$) is a tuning parameter which controls the sparsity of $\mathbf{K}$ (with larger values of λ leading to more regularization and thus more sparsity) and that is typically chosen through cross-validation. The sum $\sum_{j,k} \phi_\lambda(|\kappa_{jk}|)$ in Equation (2) can be taken with or without a penalty on the diagonal terms (Friedman et al., 2008; Meinshausen & Bühlmann, 2006; Rothman et al., 2010; Yuan & Lin, 2007).

Bayesian GLASSO

A Bayesian version of the GLASSO was introduced by Wang (2012). This Bayesian GLASSO is based on putting a double exponential prior on the off-diagonal entries of the precision matrix and an exponential prior on the diagonal entries,

$$p(\mathbf{K}|\lambda) \propto \prod_{j<k}(\mathrm{DE}(k_{jk}|\lambda)) \prod_{p=1}^{P}(\mathrm{EXP}\,(k_{jj}|\lambda/2))\mathbf{1}_{\mathbf{K}\in\mathbf{S_p}}, \tag{3}$$

where $\mathrm{DE}(x|\lambda)$ indicates the double exponential distribution with rate λ, $\mathrm{EXP}\,(x|\lambda/2)$ represents an exponential distribution with rate $\lambda/2$, and S_P is the space of $P \times P$ positive definite matrices (Li et al., 2019). The tuning parameter λ can be chosen by cross-validation as in a frequentist framework (Friedman et al., 2008; Rothman et al., 2010), or by specifying an appropriate hyperprior on this parameter (Li et al., 2019). The maximum a posteriori estimate of **K** under the prior in Equation (3) is equal to the regularized estimate of **K** you would obtain with the frequentist GLASSO (Li et al., 2019). Unlike the frequentist GLASSO, the Bayesian GLASSO does not set elements of the precision matrix to exactly 0. In fact, there is zero probability, according to the prior, that any of the partial correlations in **K** is exactly 0 (Wang, 2012). To make the Bayesian methods as comparable to the frequentist GLASSO as possible, we therefore have to apply either a discrete and continuous mixture prior distribution (Wang, 2012), such as the G-Wishart prior (Dawid & Lauritzen, 1993; Roverato, 2000), or use a heuristic decision rule to set elements of **K** to 0. In this study, we will use an additional decision rule as part of the Bayesian regularized estimation of GGMs, which sets edges whose 95% Credibility Intervals contain 0 to be equal to 0 (note that this approach is quite similar to frequentist hypothesis testing, and is only one possible decision rule). This second, decision rule (or pruning) step of our Bayesian regularization could be left out, in which case all regularization is done purely by the GLASSO or Horseshoe priors. This can be a useful alternative if having edges set to exactly to 0 is not required. However, as said above, we decided to add this second step as (an integral) part of our Bayesian regularization to keep the similarity with the frequentist GLASSO as close as possible (and to have the calculation of sensitivity, specificity, etc. make sense).

Sampling from the Bayesian GLASSO was done using the data-augmented block Gibbs sampling scheme discussed in (Wang, 2012).

Bayesian Horseshoe

An alternative penalized Bayesian Estimation method is the Graphical Horseshoe introduced by Li et al. (2019). Instead of a double exponential prior, the graphical horseshoe puts a horseshoe prior on the off-diagonal elements of **K**. The element-wise priors are specified for $(i, j = 1, \ldots, K)$ as (Li et al., 2019),

$$\begin{aligned} k_{ii} &\propto 1, \\ k_{ij:i\neq j} &\sim \mathrm{Normal}(0, \lambda_{ij}^2\tau^2), \\ \lambda_{ij:i\neq j} &\sim C^+(0, \mathbf{1}), \\ \tau &\sim C^+(0, \mathbf{1}), \end{aligned} \tag{4}$$

where C^+ denotes a half-Cauchy random variable with density $p(x) \sim (1 + x^2)^{-1}; x > 0$ (Carvalho et al., 2010; Li et al., 2019). The horseshoe prior has two shrinkage parameters; the local shrinkage parameter λ_{ij} which is unique for each unique combination of variables, and the global shrinkage parameter τ which influences all partial correlation estimates. The global shrinkage parameter adapts to the sparsity of the entire matrix **K** and shrinks the estimates of the off-diagonal elements toward zero. In contrast, the local shrinkage parameters preserve the magnitude of non-zero off-diagonal elements, and ensure that the element-wise biases are not very large (Li et al., 2019). Based on the above the horseshoe prior for **K** can be written as (Li et al., 2019),

$$p(\mathbf{K}|\tau) \sim \prod_{i<j}\mathrm{Normal}(k_{ij}|\lambda_{ij}^2, \tau^2) \prod_{i<j} C^2(\lambda_{ij}|0, \mathbf{1})\mathbf{1}_{\mathbf{K}\in S_K} \tag{5}$$

where S_K again is the space of $K \times K$ positive definite matrices (Li et al., 2019). Like the Bayesian GLASSO, the Bayesian Horseshoe does not set elements of the precision matrix to exactly 0. For that it also requires an additional decision rule like setting elements whose 95% Credibility Intervals contain 0 to be equal to 0.

Sampling from the Bayesian Horseshoe was done using the data-augmented block Gibbs sampling scheme discussed in (Li et al., 2019).

Bayesian methods for estimating centrality indices

As mentioned in the introduction, Bayesian estimation of GGMs provides posterior distributions for all sampled parameters and allows easy estimation of transformations and/or functions of these parameters. In other words, with Bayesian estimation we can combine (posteriors of) model parameters in different ways to get (posteriors of) new, composite, parameters of interest. The different centrality estimates can be

viewed as such composite parameters. We can combine (posteriors of) parameters any way we want, and as such can construct (the posterior of) a composite parameter of interest in different ways. As long as these different combinations make substantive sense, they can be viewed as representing (slightly) different operationalization of the same construct (i.e., composite parameter). As a simplistic example, if we want to construct a posterior for a composite parameter c from the posterior of model parameter a, with $c = 2 * a$, we could multiply the posterior of a by 2, or divide it by .5. Both would be different ways of getting to the posterior of c. In this study, we will test three different methods of combining (posteriors of) parameters to get estimates for the three centrality types (i.e., we will test three different Bayesian operationalizations for the centrality measures). All three methods are estimating the same constructs (the three centrality types), but do so in slightly different ways that might work more or less well in practice. Part of the aim of this study is determining which operationalization of the centrality estimates works best. The different estimation methods (operationalizations) for the three centrality types are discussed in the following three sub-sections.

Simple Gibbs-Sampler estimation

The first method used to calculate the three different centrality measures for each node involved estimating each node's strength, closeness, and betweenness centrality in each iterations of the Gibbs-sampler. Specifically, this method consists of the following steps:

Algorithm 1. Simple Gibbs-Sampler Estimation

1. For each of the k ($k = 1, \ldots, K$) iteration of the Gibbs-sampler, use the current estimate for the inverse variance-covariance matrix as input for the calculations of strength, closeness, and betweenness centrality described in Opsahl et al. (2010) and implemented in the *qgraph* R-package (Epskamp et al., 2012).
2. For each node p ($p = 1, \ldots, P$), use the K estimates of strength, closeness, and betweenness centrality obtained in the previous step to obtain the posteriors for these three centrality types for each node.
3. For each node p ($p = 1, \ldots, P$), take the modes of the posterior distributions of the centrality measures as the point estimate for the centralities of the node.
4. For each node p ($p = 1, \ldots, P$), take the 2.5th and 97.5th percentile of the posterior distributions of the centrality measures to construct the 95% Credibility intervals for each node's strength, closeness, and betweenness centrality.

A problem with this method is that, as mentioned above, the Bayesian GLASSO and Horseshoe doesn't set edges to exactly 0, which is why the Bayesian regularization employed in this study also uses a decision rule that sets all edges whose 95% Credibility Interval contains 0, to 0. This decision rule can only be applied after all iterations of the Gibbs-sampler are done however, because it requires the complete posterior distribution of the edges. Since the Simple Gibbs-Sampler Estimation approach involves calculating centrality values in each iteration of the Gibbs-sampler (so before the Gibbs-sampler is completely done), the centrality estimations mentioned in step 1 above have to be done without the (heuristic) pruning step. This will lead to positive bias in the strength and closeness centralities estimates of this method, as edges that will eventually be set to exactly 0 (and should therefore not contribute to these centrality measures at all), are not set to 0 yet and will therefore have non-zero contributions to these centrality values. Betweenness centrality could also be biased, but since this measure depends on how many shortest paths between two nodes a third node is a part of, it is less directly influenced by the values of edges. After all, regularization does not necessarily change which path between nodes is shortest. As such, the extend and form of the bias introduced in this centrality measure by the absence of the pruning step is harder to predict. Note that the edge estimates used for the calculations of the centrality values in this method are still regularized to some degree by the prior, which pulls the values toward 0. They are not *completely* regularized however, because they can't be set to exactly zero without the pruning heuristic.

To solve for this issue, and to more fully correct our centrality estimates for regularization, we also devised the 2 methods described below.

Post-processing shift estimation

The first method for calculating the centrality measures that takes complete Bayesian regularization (including pruning) into account is by using what we termed *Post-Processing Shift Estimation*. This approach consists of the following step:

Algorithm 2 Post-Processing Shift Estimation

1. Calculate the posteriors for centrality measures according to the Simple Gibbs-Sampler Estimation method discussed above.
2. Prune the final estimate of the precision matrix by setting edges whose 95% CI contains 0 to 0. Point estimates of the edges are based on the modes of the posterior (MAP estimates).
3. For each node p ($p = 1, \ldots, P$), estimate the three centrality values based on the pruned precision matrix from step 2 (using the calculations described in Opsahl et al. (2010) and implemented in the *qgraph* R-package (Epskamp et al., 2012)).
4. For each node p ($p = 1, \ldots, P$), calculate the difference between the point estimates of the three centrality measures obtained with Simple Gibbs-Sampler Estimation method (in step 3) and the point estimates obtained in the previous step. The difference between these two estimates can be interpreted as the bias introduced in the centrality estimates by not setting edges to exactly 0.
5. For each node p ($p = 1, \ldots, P$), use the difference between the two sets of point estimates calculated in the previous step to shift the posteriors for the three centrality measures obtained with the Simple Gibbs-Sampler Estimation method so that the modes of these posteriors become equal to the point estimates for the centralities obtained from the pruned precision matrix (step 3).
6. For each node p ($p = 1, \ldots, P$), take the 2.5th and 97.5th percentiles of the shifted posteriors from the previous step to construct the 95% Credibility intervals for each node's strength, closeness, and betweenness centrality.

Note that since we're merely shifting the posteriors obtained from the Simple Gibbs-Sampler Estimation method to construct our new posteriors with this method, the widths of the intervals will be the same across these two methods. However, in this method the posteriors are centered around less biased point estimates of strength, closeness, and betweenness centrality.

For the Simple Gibbs-Sampler Estimation method, the posterior distributions of the centrality measures were restricted to be larger than or equal to 0, as they should be, since the values of the centrality measures can't be smaller than 0. By shifting these posteriors, as we do in this second method, we could make some of the posteriors cover values smaller than 0 as well. In practice however, this will likely not cause any problems. First, in the tests analyses that we ran, the 95% Credibility Intervals for the centrality measures of the Post-Processing Shift method never went below 0. In addition, even if they did, this would just imply that we can't rule out that the corresponding node is unconnected to all others in the network (although the negative values should obviously not be interpreted).

Estimation based on edge weight estimates

The second fully regularized estimation method for the centrality measures is not based on calculating the centrality values for each node in each iteration of the Gibbs-sampler. Instead it is based on the estimates of the edges themselves and the mathematical expressions for the variance of a(n) (inverse of a) sum.

For random variables $Y_1, \ldots, Y_k$, the variance of their sum is given by,

$$var\left[\sum_{k=1}^{K} Y_k\right] = \sum_{k=1}^{K} \sigma_k^2 + 2 \sum_{1 \le k \le l \le K} \sigma_k \sigma_l, \qquad (6)$$

where σ_k^2 is the variance of variable k. In addition, the variance of the inverse of a random variable Y is given by Mood et al. (1985),

$$var\left[\frac{1}{Y}\right] \approx \left(\frac{1}{\mu_Y}\right)^2 \left(\frac{\sigma_Y^2}{\mu_Y^2}\right) \qquad (7)$$

This second Equation can be obtained by a Taylor-series expansion and dropping all terms of order higher than 2 (Mood et al., 1985).

For strength centrality, the steps of this Estimation Based approach are as follows:

Algorithm 3 Estimation Based Strength Centrality

1. Prune the final estimate of the precision matrix by setting edges whose 95% CI contains 0 to 0. Point estimates of the edges are based on the modes of the posterior (MAP estimates).
2. For each node p ($p = 1, \ldots, P$), estimate the strength centrality based on the pruned precision matrix from the previous step (using the calculations described in Opsahl et al. (2010)).
3. For each edge k ($k = 1, \ldots, K$), that is, each element of the inverse covariance matrix, use it's posterior to determine its variance σ_k^2.
4. For each pair of edges k and l ($k = 1, \ldots, K$, $l = 1, \ldots, K$, $k \neq l$), use their values from each iteration of the Gibbs-Sampler to calculate the covariance between the two σ_{kl}.
5. For each node p ($p = 1, \ldots, P$), calculate the variance of the sum of it's edges (to all other nodes) using Equation (6): $\sigma^2_{Strength,p} = \sum_{k=1}^{S} \sigma_k^2 + 2 \sum_{1 \le k \le l \le S} \sigma_{kl}$ (for $k = 1, \ldots, S$, $l = 1, \ldots, S$, $k \neq l$), where σ_k^2 is the variance of edge k, σ_{kl} is the covariance between edge k and edge l, and S is the maximum number of edges of node p. Note that we also take the variance of the edges that are set to 0 after regularization into account when calculating the variance of the strength centrality of a node.
6. For each node p ($p = 1, \ldots, P$), calculate the 95% Credibility Intervals for each node's strength centrality by taking the point estimate for the strength centrality of that node (step 2) and adding and subtracting $1.96 * \sigma^2_{Strength,p}$ to the point estimate, where $\sigma^2_{Strength,p}$ is estimated in the previous step.

For closeness centrality, the steps of the Estimation Based approach are:

Algorithm 4 Estimation Based Closeness Centrality

1. Prune the final estimate of the precision matrix by setting edges whose 95% CI contains 0 to 0. Point estimates of the edges are based on the modes of the posterior (MAP estimates).
2. For each node p ($p = 1, \ldots, P$), estimate the closeness centrality based on the pruned precision matrix from the previous step (using the calculations described in Opsahl et al. (2010) and implemented in the *qgraph* R-package (Epskamp et al., 2012)).
3. For each edge k ($k = 1, \ldots, K$)(i.e., element of the inverse covariance matrix), use it's posterior to determine its variance σ_k^2.
4. For each pair of edges k and l ($k = 1, \ldots, K$, $l = 1, \ldots, K$, $k \neq l$), use their values from each iteration of the Gibbs-Sampler to calculate the covariance between the two σ_{kl}.
5. For each node p ($p = 1, \ldots, P$), determine the edges on the largest/strongest shortest path from the node to all other nodes using Dijkstra's algorithm (Dijkstra, 1959) (implemented in the *qgraph* R-package (Epskamp et al., 2012))).
6. For each pair of nodes p and m ($p = 1, \ldots, P, m = 1, \ldots, P, p \neq m$), calculate the variance of the sum of the edges of the shortest path connecting the two nodes (determined in the previous step) using Equation (6): $\sigma^2_{\text{Shortest Path},pm} = \sum_{k=1}^{S} \sigma_k^2 + 2\sum_{1 \leq k \leq l \leq S} \sigma_{kl}$, where σ_k^2 is the variance of edge k, σ_{kl} is the covariance between edges k and l, and S is the maximum number of edges on the shortest path between the two nodes.
7. For each node p ($p = 1, \ldots, P$), calculate the variance of the sum of all of its shortest paths to all other nodes using Equation (6): $\sigma^2_{\text{Sum of Shortest Paths of node p}} = \sum_{m=1}^{P} \sigma^2_{\text{Shortest Path},pm} + 2\sum_{1 \leq m \leq n \leq P} \sigma_{pm,pn}$ (for $m \neq p$), where $\sigma^2_{\text{Shortest Path},pm}$ is the variance of the shortest path between nodes p and m (determined in the previous step), and $\sigma_{pm,pn}$ is the covariance between the shortest path from node p to node m, and the shortest path from node p to node n.
8. For each node p ($p = 1, \ldots, P$), calculate the mean of the sum of all of its shortest paths to all other nodes. This is done by calculating the sum of the shortest paths between the node and all other nodes in each iteration of the Gibbs-sampler, and then taking averaging across iterations.
9. For each node p ($p = 1, \ldots, P$), calculate the variance of the closeness centrality using Equation (7): $\sigma^2_{Clsnss,p} = \frac{\sigma^2_{\text{Sum of Shortest Paths of node p}}}{\mu^4_{\text{Sum of Shortest Paths of node p}}}$, where $\mu_{\text{Sum of Shortest Paths of node p}}$ is the mean of the sum of all of shortest paths from node p to all other nodes (calculated in previous step).
10. For each node p ($p = 1, \ldots, P$), calculate the 95% Credibility Intervals for each node's closeness centrality by taking the point estimate for the closeness centrality of that node (step 2) and adding and subtraction $1.96 * \sigma^2_{Clsnss,p}$ to the point estimate, where $\sigma^2_{Clsnss,p}$ is estimated in the previous step.

Note that in the steps above we have two "types" of paths when combining variances and covariances into total variance. The first type are the successive edges on the same shortest path (e.g., on the path A – B – C, we have the edges/paths A – B and B – C), and the second are the separate shortest paths emanating from the same node (e.g., the paths A – B – C and A – D – E). We aren't sure whether it makes sense for these two types of paths to be correlated amongst each other (i.e., for successive edges on the same shortest path to be correlated, or for separate shortest paths emanating from the same node to be correlated). As a result we used four different variations of our edge weight based calculation of closeness centrality. We calculated the total closeness centrality assuming that 1) both successive edges on the same shortest path and different shortest paths emanating from the same node could be correlated, 2) only separate shortest paths emanating from the same node could be correlated, 3) only successive edges on the same shortest paths could be correlated, and 4) assuming that both shortest and successive paths were uncorrelated.

A downside of this second, edge weight based, method is that it can't be used for a new estimate of betweenness centrality, as it isn't a direct function of individual edges weights. Instead, betweenness centrality is about how many shortest paths between two other nodes a node is part of. In addition, it assumes normally distributed random variables. This second downside, isn't necessarily a large problem however. In our data, the posterior distributions of most edges look pretty normal. Furthermore, the width of the CIs for strength centrality of this method also appear quite similar to those obtained with the other two methods (that don't assume normality). Sometimes they are a little wider, sometimes a little narrower but the difference is always in the second decimal place or even smaller.The width of CIs for closeness centrality are also quite similar. They tend to be a little narrower of wider depending on whether one assumes successive edges on shortest paths to be correlated (if successive edges on a shortest path are assumed to be correlated the intervals tend to be a little wider than those of the other methods, while assuming independent successive edges on a shortest path leads to slightly narrower intervals than the other methods), but these differences tend to be in the 4 decimal place or even smaller (typically the fifth decimal place).

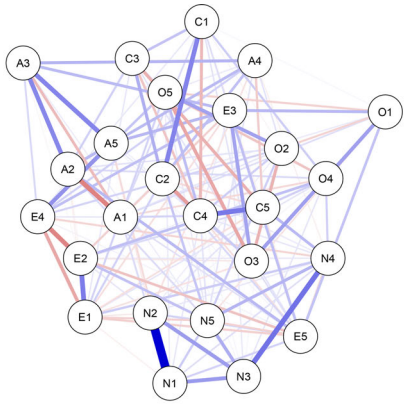
Simulation study

To test the performance of the two Bayesian estimation methods mentioned above and the different centrality estimation methods we ran an extensive simulation study. Our focus for this paper is on the social sciences (specifically psychology), which is why we generated data that is representative for data encountered in

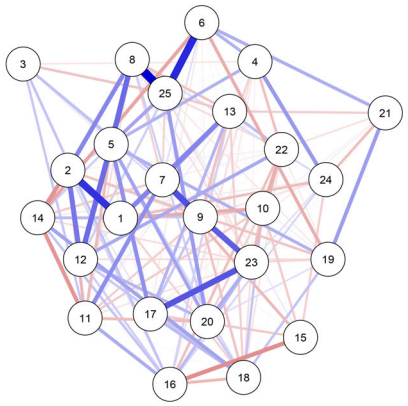
this field (while also looking at the effect of network structure and sparsity on performance). Specifically, we generated data for two different network structures: 1) a network based on the partial correlation from the bfi-data in the *Psych* R-package (Revelle, 2019), which contains information on 25 personality self-report items taken from the International Personality Item Pool (ipip.ori.org), and 2) a random network structure for 25 nodes generated with the *genGGM* function from the *bootnet* package (Epskamp et al., 2017) (Figure 2). We generated data for sample sizes of $N = 900$, 5,000 and 10,000 (which gives us N (sample size) to p (number of estimated parameters) ratios of 3, 16.67, and 33.33 respectively). In addition, we also varied the density of the random network so that it either closely matched that of the bfi-data (.537 vs .517) or was equal to about half that of the bfi-data (.263 vs .517). The partial correlations in the bfi-data ranged from −.256 to.528 (mean=.023, sd=.080), the partial correlations in the random network of equal density ranged from −.216 to.492 (mean=.013, sd=.093), and the partial correlations in the random network of half-density ranged from −.319 to.510 (mean=–.001, sd=.089). For each scenario (i.e., combination of N and network-structure) we ran 1000 replications.

For the Bayesian estimation of the generated data we used the Bayesian GLASSO prior (with $\lambda \sim gamma(1, .01)$, i.e. we used a gamma distribution with rate parameter of 1 and scale parameter of .01 as hyperprior for the tuning parameter) and Bayesian Horseshoe prior (with $C^{+}(0, \mathbf{1})$ hyperpriors for both the local (λ_{ij}) and global (τ) shrinkage parameters) discussed above. For both methods we used 10,000 burn-in iterations and 10,000 subsequent iterations. Traceplots showed good convergence with these number of iterations. We also analyzed the generated data using a frequentist graphical LASSO for comparison. For this frequentist estimation we used the estimation method used in the popular *bootnet* R-package (Epskamp et al., 2017), in which the optimal value for the regularization parameter was selected using the Extended Bayesian Information Criterion (EBIC). Specifically, this frequentist method runs the Glasso estimation 100 times (with 100 different values for the tuning parameter). The values of the tuning parameter in these runs are logarithmically spaced between the maximal value of the tuning parameter at which all edges are zero (λ_{max}) and $\lambda_{max}/100$. For each of the resulting graphs the EBIC is computed and the graph with the best EBIC is selected (Epskamp et al., 2012).

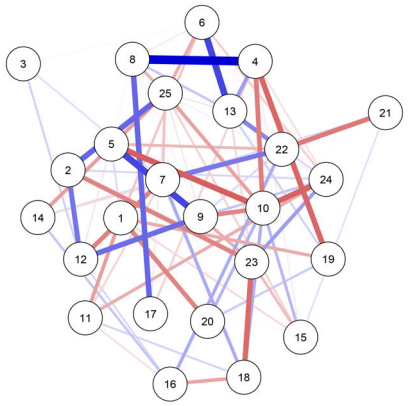
To assess performance, we first compared the performance of the Bayesian GLASSO and Horseshoe to frequentist GLASSO estimation on KL-loss, the Frobenius norm, the correlation between the estimated and true precision-matrix elements, F1 (the harmonic mean of the positive predictive value and sensitivity), sensitivity, and specificity. The KL-loss and Frobenius norm are measures of the "distance" between the true and estimated precision matrix and can therefore be viewed as measures of bias in edge estimates. The correlation between the estimated and true precision-matrix



a Psych Network: Network Based on the bfi-data

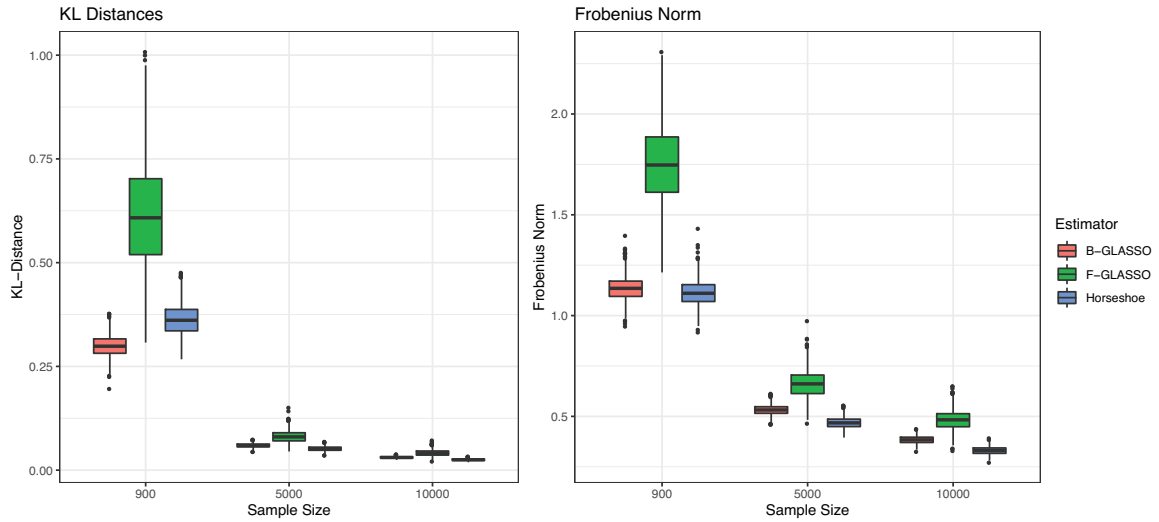


b RED Network: Random Network with density equal to the Psych Network

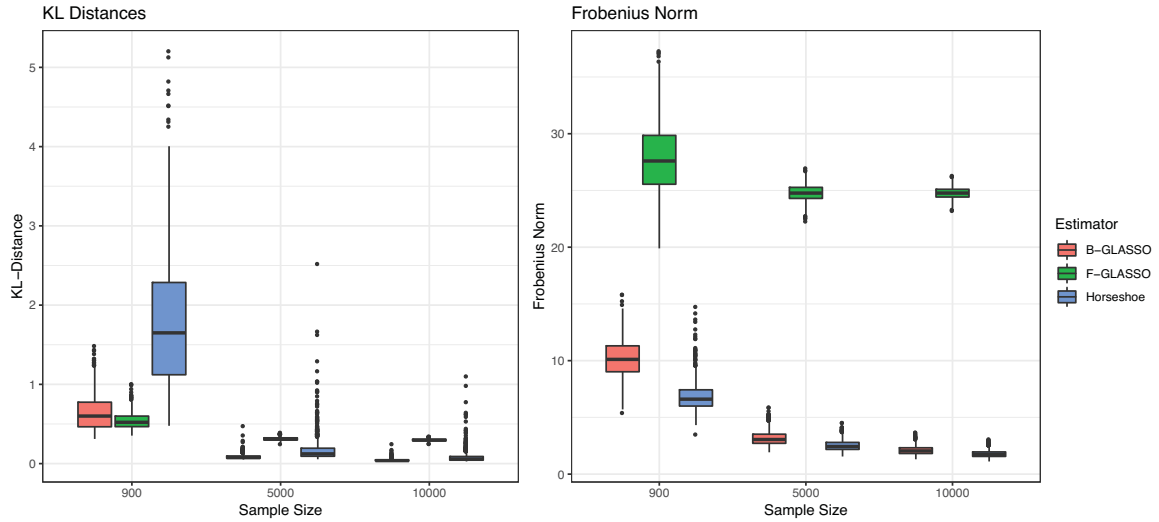


c RHD Network: Random Network with density half that of the Psych Network

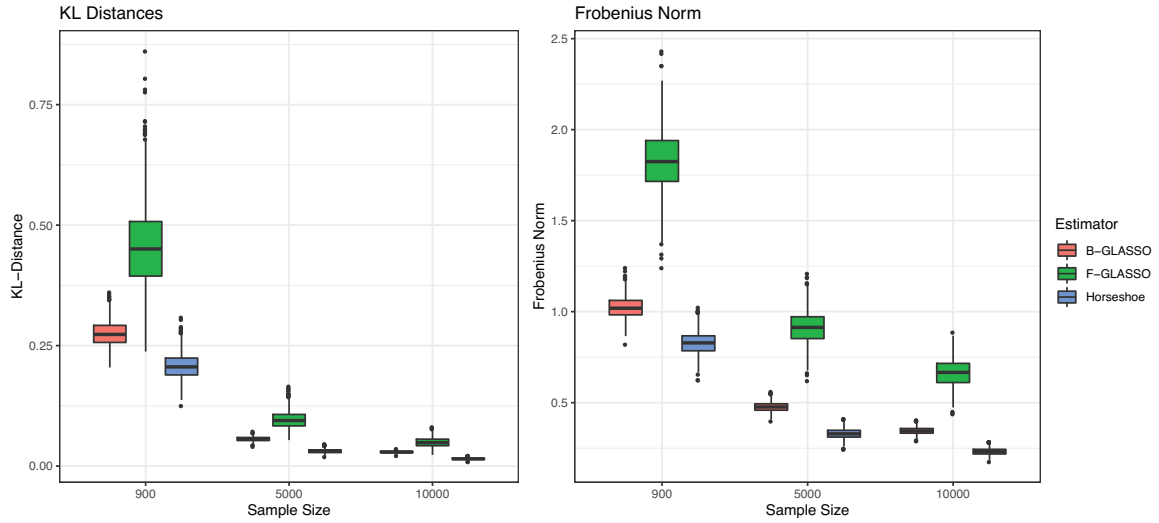
Figure 2. The three network-structures under which data were generated. (a) Bias for Psych Network. (b) Bias for Random Network. (c) Bias for Random Network (Half Density).



a Bias for Psych Network



b Bias for Random Network



c Bias for Random Network (Half Density)

Figure 3. Bias in estimated edge weights for the different estimation methods at N = 900, 5,000, and 10,000.

elements give an indication of how well the rank-order of edges are maintained. The sensitivity and specificity give an indication of how well edges are classified as present or absent respectively, while F1 gives and indication of how well edges that are present in the population are identified as such.

Next, we assessed performance of the Bayesian methods with regard to the different centrality measures by looking at the mean square error of the centrality estimates, the correlation between the true and estimated centrality measures, and the coverage of the 95% Credibility Intervals (both per node and on average). Coverages between .92 and .98 were deemed acceptable (Bradley, 1978), and per node coverage was deemed sufficient if at least 75% of nodes had acceptable coverage.

Results

Comparison to frequentist GLASSO

Figure 3 shows that the Bayesian estimation methods compare well to the frequentist GLASSO in terms of total bias in the (regularized) partial correlation matrix. For both the network based on the bfi-data from the *Psych* R-package (the Psych Network) (Revelle, 2019) and the Random Network with half the density of the Psych Network (the Random Half Density or *RHD* Network) the two Bayesian methods show lower KL-distance and lower Frobenius norms than the frequentist GLASSO for all three sample sizes. For the Random Network with equal density (the Random Equal Density or *RED* Network), the frequentist GLASSO performs better that the Horseshoe, and slightly better than the Bayesian GLASSO, with respect to KL-distance at $N = 900$. For sample size of 5,000 and 10,000, the two Bayesian methods perform better however. The Bayesian GLASSO and Horseshoe also perform better in terms of the Frobenius Norm for the RED Network at all sample sizes.

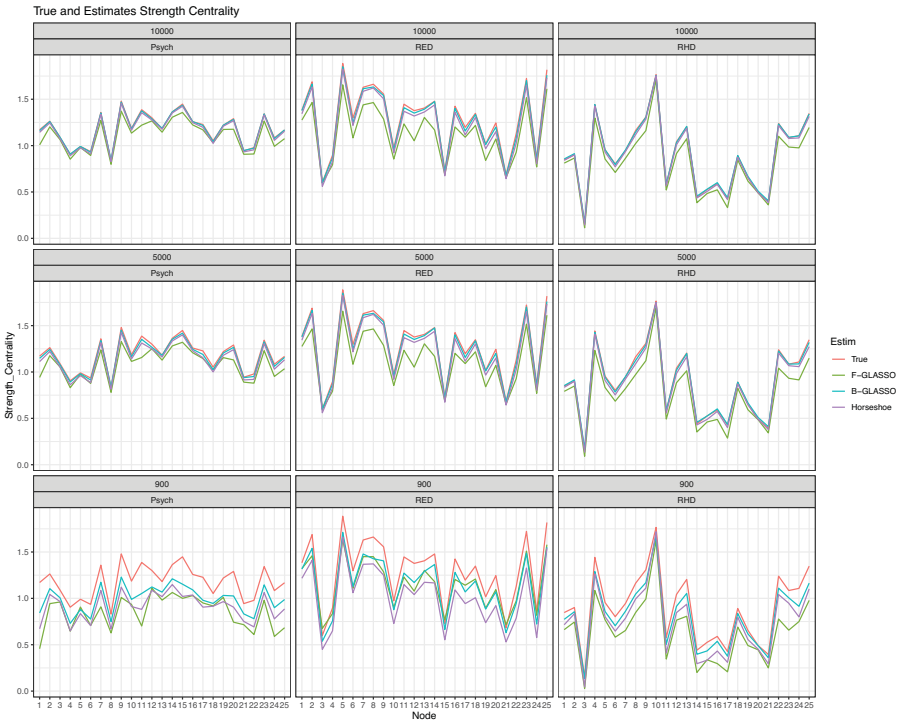
To improve readability, figures for the the correlation between the estimated and true precision-matrix elements, sensitivity, and specificity are given in Appendix A. Figure A1 shows that all three estimation methods show high correlations between true and estimated edge weights at all sample sizes (with correlations becoming larger as N increases). Again the Bayesian method appear to have a slight edge over the frequentist GLASSO. For all networks, the Bayesian GLASSO always has higher correlations than its frequentist counterpart, while the Horseshoe has higher correlations for $N > 900$.

Figure A2 shows that for the Psych Network, the frequentist GLASSO has better sensitivity that the Bayesian methods at all sample sizes. For the RED Network (Figure A3) the frequentist GLASSO works better for $N < 10,000$. At $N = 10,000$ the Bayesian GLASSO performs best (although the difference with the frequentist GLASSO is not large). For the RHD Network (Figure A4), the frequentist GLASSO always has better sensitivity than the Graphical Horseshoe. At $N > 900$, however, the sensitivity of the Bayesian GLASSO is equal or slightly better. The specificity is always better for the Bayesian methods, while F1 is better for all networks for sample size larger than 900. For the two random networks the F1 for the Bayesian GLASSO is also larger than that of the frequentist GLASSO at $N = 900$.

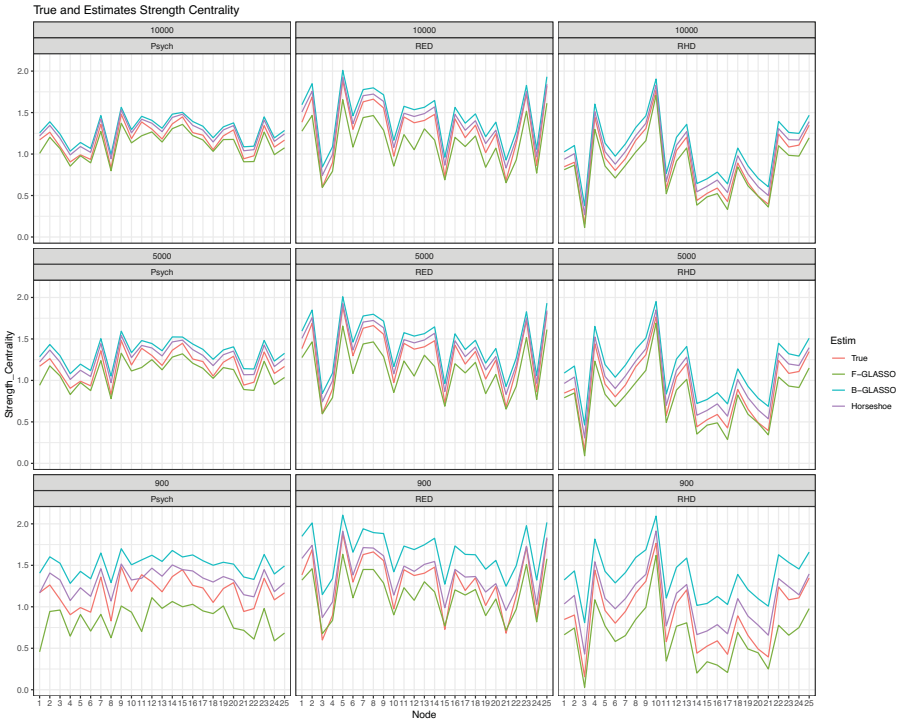
Figures A5–A7 show the correlation between the true and estimated centrality measures. For the Psych Network (Figure A5), the correlation between true and estimated strength and closeness centrality is always higher for the Bayesian methods (although the frequentist GLASSO also has high correlations). The correlation between true and estimated betweenness centrality is about the same for all estimation methods. For the RED network (Figure A6), the Bayesian methods have higher correlations between true and estimated closeness and betweenness centrality. The correlation between true and estimated strength centrality is better for this network if $N > 900$. All three measures are always really close in performance. Finally, for the RHD network (Figure A7), the Bayesian GLASSO always has higher correlation between true and estimated strength and closeness centrality, and has the highest correlations for betweenness centrality if $N > 900$ (otherwise the variability in correlation is slightly larger than for the other methods). The Horseshoe has higher correlations than the frequentist GLASSO for strength and betweenness centrality, but the correlation for closeness centrality in this network is really unstable for this estimation method.

Centrality measure coverage

For the Psych, RED, and RHD networks, Figures 4a,b, 6a,b, and 8a,b show the true Strength, Closeness, and Betweenness centrality of each of the 25 nodes respectively (in red). In addition, they show the estimated values of these centralities for the Frequentist GLASSO (in green), the Bayesian GLASSO (in blue), and the Graphical Horseshoe (in purple), for each of the three sample sizes under which data was generated



a True Strength Centrality for Post-Processing Shift Estimation



b True and Estimated Strength Centrality for Simple Gibbs-Sampler

Figure 4. True and estimated strength centrality.

from the networks. For the Bayesian methods, estimated obtained with the Post-Processing Shift estimation method are given in Figures 4a, 6a, and 8a, while the estimated obtained with the Simple Gibbs-Sampler estimation method are given in Figures 4b, 6b, and 8b.

For the Bayesian estimation methods, Figures 5, 7, and 9 give information on the average coverage rates

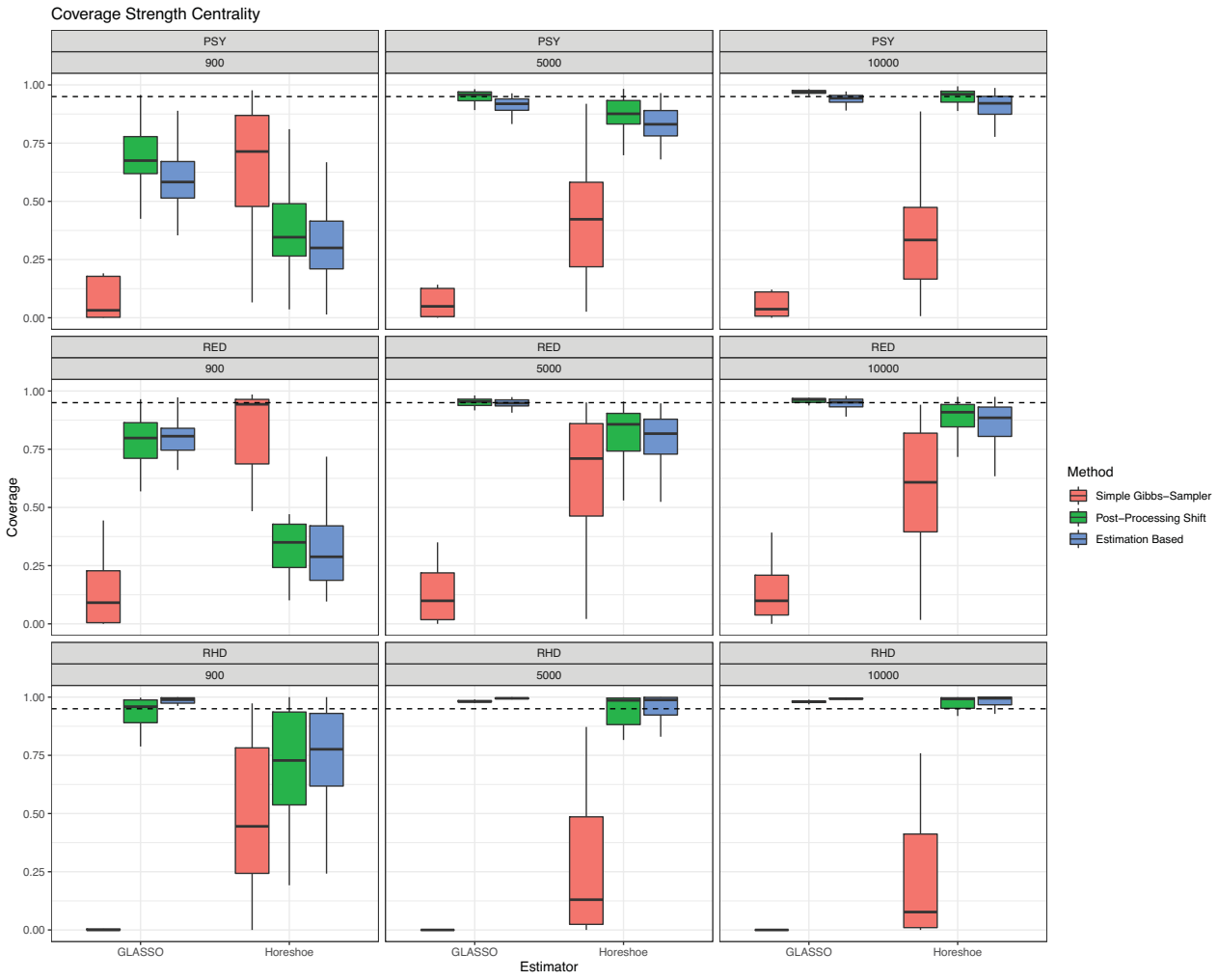


Figure 5. Coverage for strength centrality.

and the variability in these rates across the 25 nodes (for Strength, Closeness, and Betweenness centrality respectively), for each combination of the three networks and the three generated sample sizes.

For readability, tables with the detailed centrality information for each of the 25 nodes in each of the three networks are given in an online Appendix (https://osf.io/9kjxv/files/).

Strength

When using the Bayesian GLASSO, strength centrality of all networks is estimated best by the Post-Processing Shift estimation method. This method has the lowest MSE (tied with the estimation method based on edge weights which uses the same point estimate) (see Figures 4a–4b and the Tables A.1–A.3 in the online Appendix).

Importantly, for $N > 900$, this method also has good coverage for the two denser networks (the Psych- and RED Network) (Figure 5 and Table 1). For the Psych Network, the average coverage rate and proportion of nodes with sufficient coverage is .949 and .760 for $N = 5,000$, and .969 and .920 for $N = 10,000$. For the RED network the average coverage and proportion of nodes with sufficient coverage is .941 and .760 for $N = 5,000$ and .951 and .800 for $N = 10,000$. For both these denser networks coverage at $N = 900$ is too low due to negative bias in the point estimates for each node strength centrality. For the less dense RHD network, the Post-Processing Shift estimation method has good average coverage at all sample sizes, but the proportion of nodes with sufficient coverage is too low, although, at $N > 900$, this is due to the method being too conservative (i.e., to many coverages above .98 (see Table A.3 of the online Appendix)).

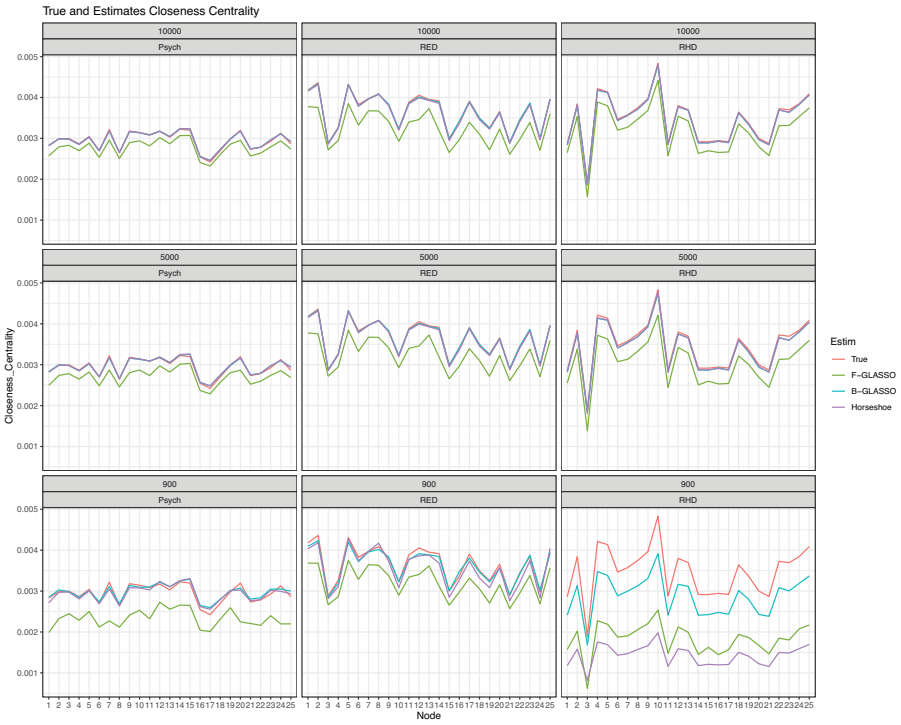
Estimation using the Simple Gibbs-sampler estimation method resulted in positive bias in the strength estimates as expected, and as a result average and per node coverage where too low for all networks and sample sizes.

Table 1. Average coverage rates for the Bayesian GLASSO.

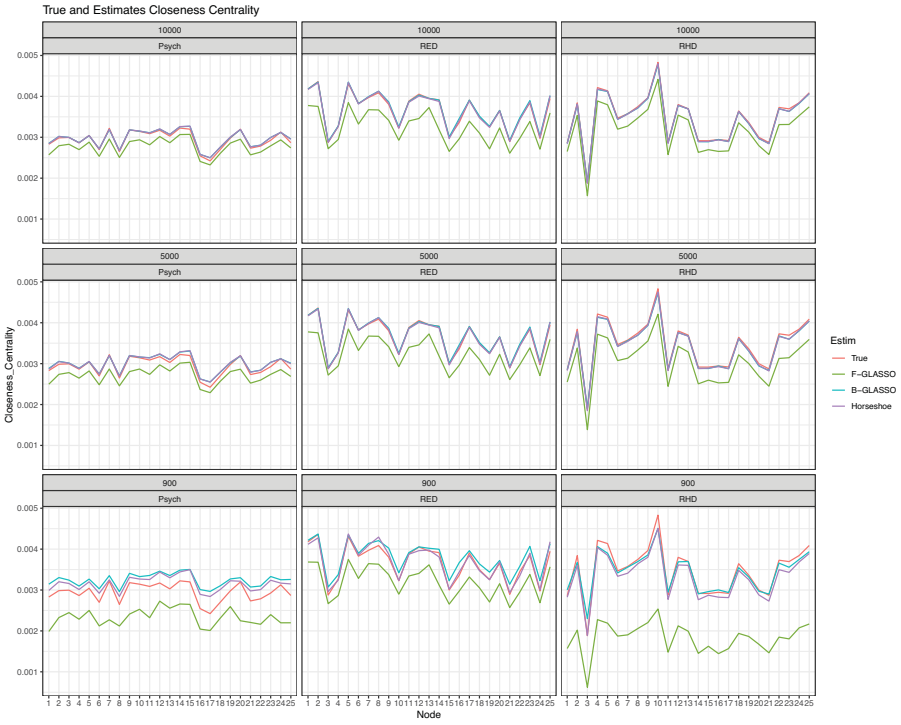
Centrality	Estimation method	Psych Network						Random network (Eq. density)						Random network (half density)					
		Sample size						Sample size						Sample size					
		900		5,000		10,000		900		5,000		10,000		900		5,000		10,000	
		Average	%	Average	%	Average	%	Average	%	Average	%	Average	%	Average	%	Average	%	Average	%
Strength	Post-processing shift	.661	.120	.949	.760	.969	.920	.775	.160	.941	.760	.951	.800	.923	.400	.978	.440	.980	.480
	Simple Gibbs-sampler	.152	.000	.135	.000	.124	.000	.139	.000	.141	.000	.148	.000	.008	.000	.005	.000	.004	.000
	Based on edge weights	.583	.040	.910	.480	.939	.840	.776	.160	.935	.800	.943	.840	.967	.200	.993	.080	.993	.000
Closeness	Post-processing shift	.932	.720	.938	.880	.940	.880	.939	.840	.948	.960	.950	1.000	.687	.000	.888	.240	.920	.600
	Simple Gibbs-sampler	.564	.120	.899	.600	.923	.720	.908	.320	.952	.920	.956	.960	.926	.440	.912	.680	.932	.720
	Based on edge weights:																		
	Correlated successive and shortest paths	1.000	.000	.979	.520	.964	1.000	1.000	.000	.970	.920	.959	1.000	.955	.680	.935	.720	.930	.720
	Correlated successive paths	1.000	.000	.913	.560	.871	.120	1.000	.000	.884	.320	.854	.160	.949	.720	.769	.000	.749	.000
	Correlated shortest paths	1.000	.000	.979	.480	.964	1.000	1.000	.000	.970	.840	.959	1.000	.955	.680	.935	.720	.930	.720
	Uncorrelated successive and shortest paths	1.000	.000	.914	.600	.871	.120	1.000	.000	.885	.680	.854	.160	.949	.720	.769	.000	.749	.000
Betweenness	Post-processing shift	.806	.000	.891	.600	.905	.720	.845	.240	.903	.640	.917	.640	.857	.400	.953	.640	.955	.600
	Simple Gibbs-sampler	.956	.440	.973	.520	.979	.400	.963	.320	.973	.320	.980	.200	.932	.240	.991	.200	.990	.200

Table 2. Average coverage rates for the Bayesian horseshoe.

Centrality	Estimation method	Psych Network						Random network (Eq. density)						Random network (half density)					
		Sample size						Sample size						Sample size					
		900		5,000		10,000		900		5,000		10,000		900		5,000		10,000	
		Average	%	Average	%	Average	%	Average	%	Average	%	Average	%	Average	%	Average	%	Average	%
Strength	Post-processing shift	.389	.000	.876	.280	.950	.600	.373	.000	.813	.160	.874	.440	.703	.160	.933	.160	.966	.320
	Simple Gibbs-sampler	.646	.200	.432	.000	.360	.000	.773	.440	.610	.080	.148	.040	.482	.040	.260	.000	.198	.000
	Based on edge weights	.336	.000	.818	.160	.911	.440	.350	.000	.783	.120	.849	.320	.742	.080	.946	.020	.975	.240
Closeness	Post-processing shift	.921	.680	.937	.800	.941	.800	.882	.240	.943	.840	.950	.960	.344	.000	.892	.280	.931	.720
	Simple Gibbs-sampler	.752	.240	.898	.560	.924	.760	.952	.680	.962	.960	.965	.960	.848	.360	.917	.640	.943	.880
	Based on edge weights:																		
	Correlated successive and shortest paths	1.000	.000	.978	.400	.964	.920	1.000	.000	.968	.800	.960	.840	.851	.320	.958	.840	.941	.880
	Correlated successive paths	1.000	.000	.912	.440	.869	.120	1.000	.000	.878	.320	.856	.200	.839	.240	.825	.160	.780	.000
	Correlated shortest paths	1.000	.000	.978	.400	.964	.920	1.000	.000	.968	.800	.960	.840	.851	.320	.958	.840	.941	.880
	Uncorrelated successive and shortest paths	1.000	.000	.912	.480	.870	.120	1.000	.000	.878	.320	.856	.200	.839	.240	.826	.160	.780	.000
Betweenness	Post-processing shift	.777	.000	.895	.600	.907	.760	.819	.320	.897	.640	.918	.600	.893	.480	.952	.600	.958	.560
	Simple Gibbs-sampler	.967	.360	.975	.560	.982	.360	.960	.280	.973	.360	.981	.200	.978	.120	.991	.160	.991	.160



a True and Estimated Closeness Centrality for Post-Processing Shift Estimation



b True and Estimated Closeness Centrality for Simple Gibbs-Sampler

Figure 6. True and estimated closeness centrality.

Estimation based on edge weights performed well in terms of MSE, but performed worse than the Post-Processing Shift approach in terms of coverage. This method only showed good coverage for $N = 10{,}000$ in the Psych Network, and $N > 900$ in the RED network.

When using the graphical Horseshoe, strength centrality is estimated less well than when the Bayesian GLASSO is used (see Table 2, Tables A.4–A.6 of the online Appendix, and Figures 4a and 4b). The MSE's of both the Post-Processing Shift estimation method

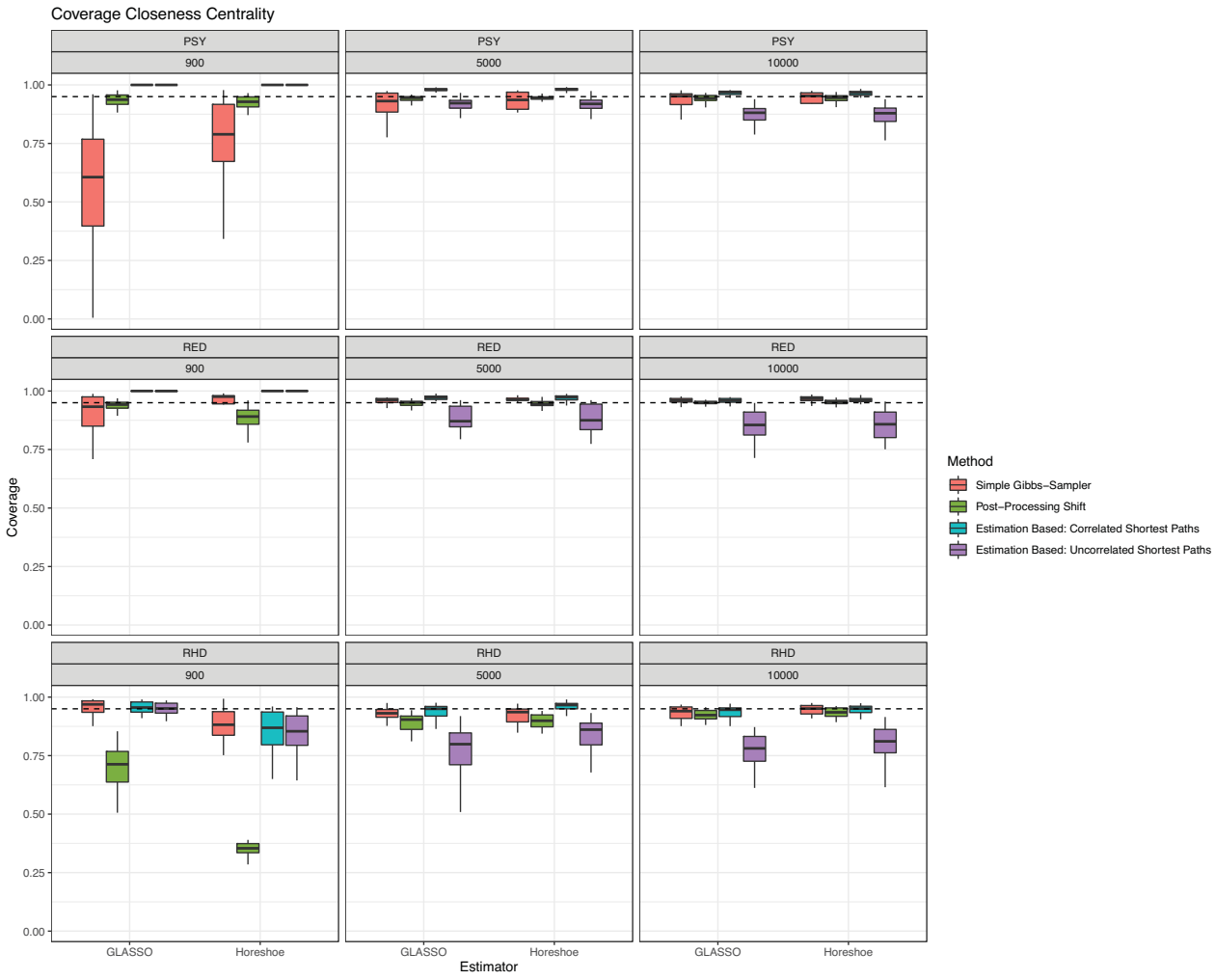


Figure 7. Coverage for closeness centrality.

and the estimation based on edge weights are slightly larger (although the difference is very small) while the average and per node coverage of these methods is insufficient for all networks and sample sizes. Simple Gibbs-sampler estimation method does have smaller MSE's when the Horseshoe is used, but as mentioned above, there is substantial positive bias, and therefore, low coverage when using this method.

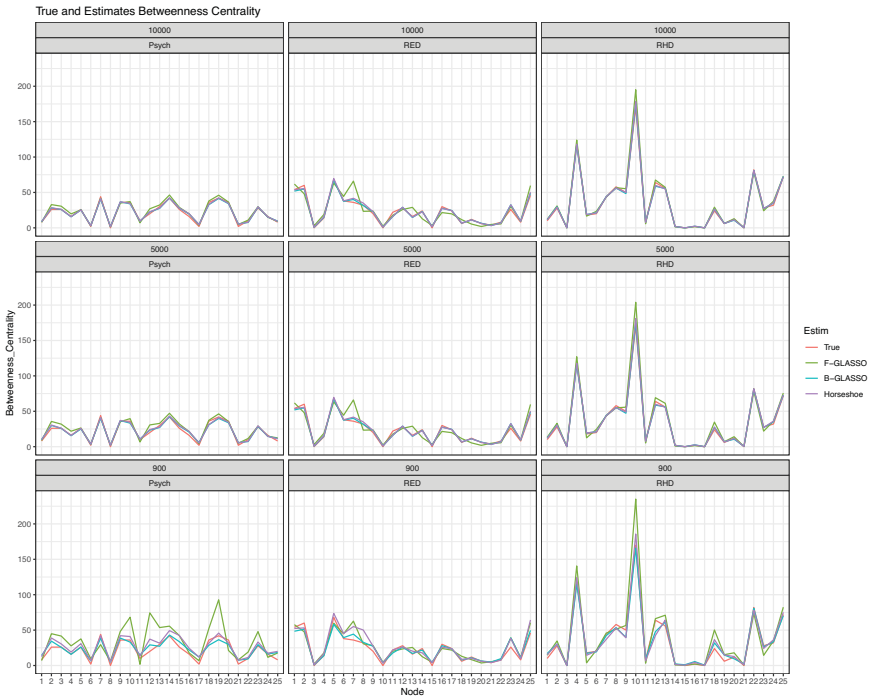
Closeness

When using the Bayesian GLASSO, closeness centrality of the denser Psych- and RED networks is again estimated best by the Post-Processing Shift Estimation method. This method had low MSE and good average coverage for all N (see Table 1, Tables A.1–A.3 of the online Appendix, and Figures 6a and 6b).

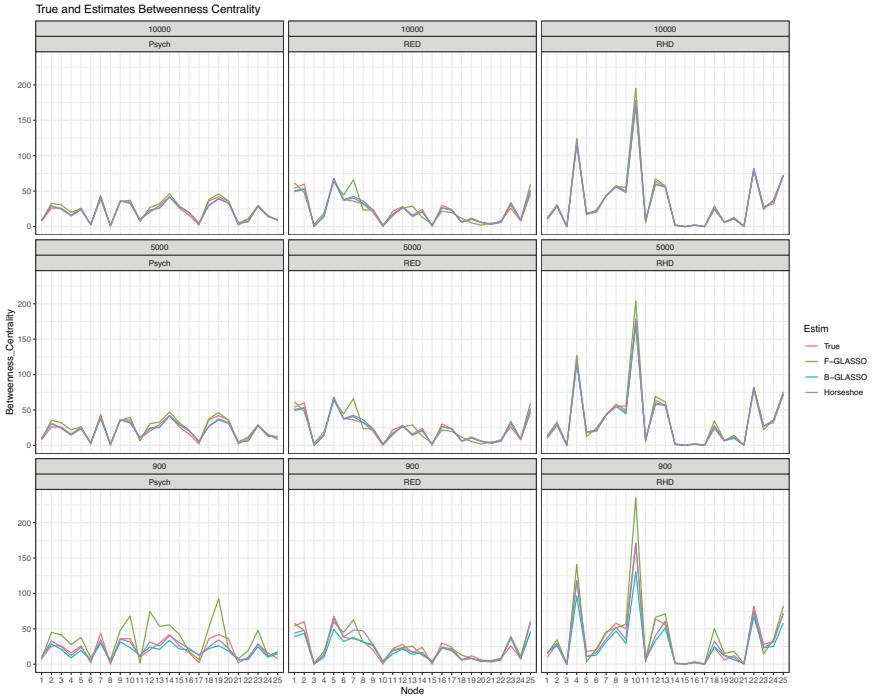
The proportion of nodes with sufficient coverage is also good for this method in these denser networks (Table 1), except for $N = 900$ in the Psych Network where it is a little too low (.720). For the sparser RHD network, the coverage of closeness centrality of this method is too low.

Estimation using the Simple Gibbs-Sampler estimation method resulted in positive bias in the closeness estimates, as expected, for the denser Psych and RED networks. For the RHD network, there was a small negative bias (except at $N = 900$). This is probably due to the fact that there are a lot more edges that were close to 0 (in each iteration of the Gibbs-sampler) in this network, which could have limited the effect of the Bayesian GLASSO not setting edges explicitly to 0 when estimating a network. Average and per node coverage where too low for this method however, except for N of 5,000 or 10,000 in the RED network.

Of the 4 estimation methods based on edge weights, the methods with correlated successive and shortest paths and the method with correlated shortest paths worked best. In fact, performance of these two methods was exactly the same, as was that of the two methods assuming either uncorrelated shortest paths or uncorrelated successive and shortest paths. This



a True and Estimated Betweenness Centrality for Post-Processing Shift Estimation



b True and Estimated Betweenness Centrality for Simple Gibbs-Sampler

Figure 8. True and estimated betweenness centrality.

indicates that successive paths on a shortest route between nodes are (effectively) uncorrelated. In the rest of this paper we will therefore only distinguish between methods with correlated shortest paths and the methods with uncorrelated shortest paths. The method with correlated successive paths showed good coverage for $N = 10{,}000$ in the Psych Network, and for $N > 900$ in the RED network. For lower sample sizes, this method was too conservative in these denser networks. In the sparser RHD network, this method did not have good coverage. This could be due to shrinkage being "stronger" in this network

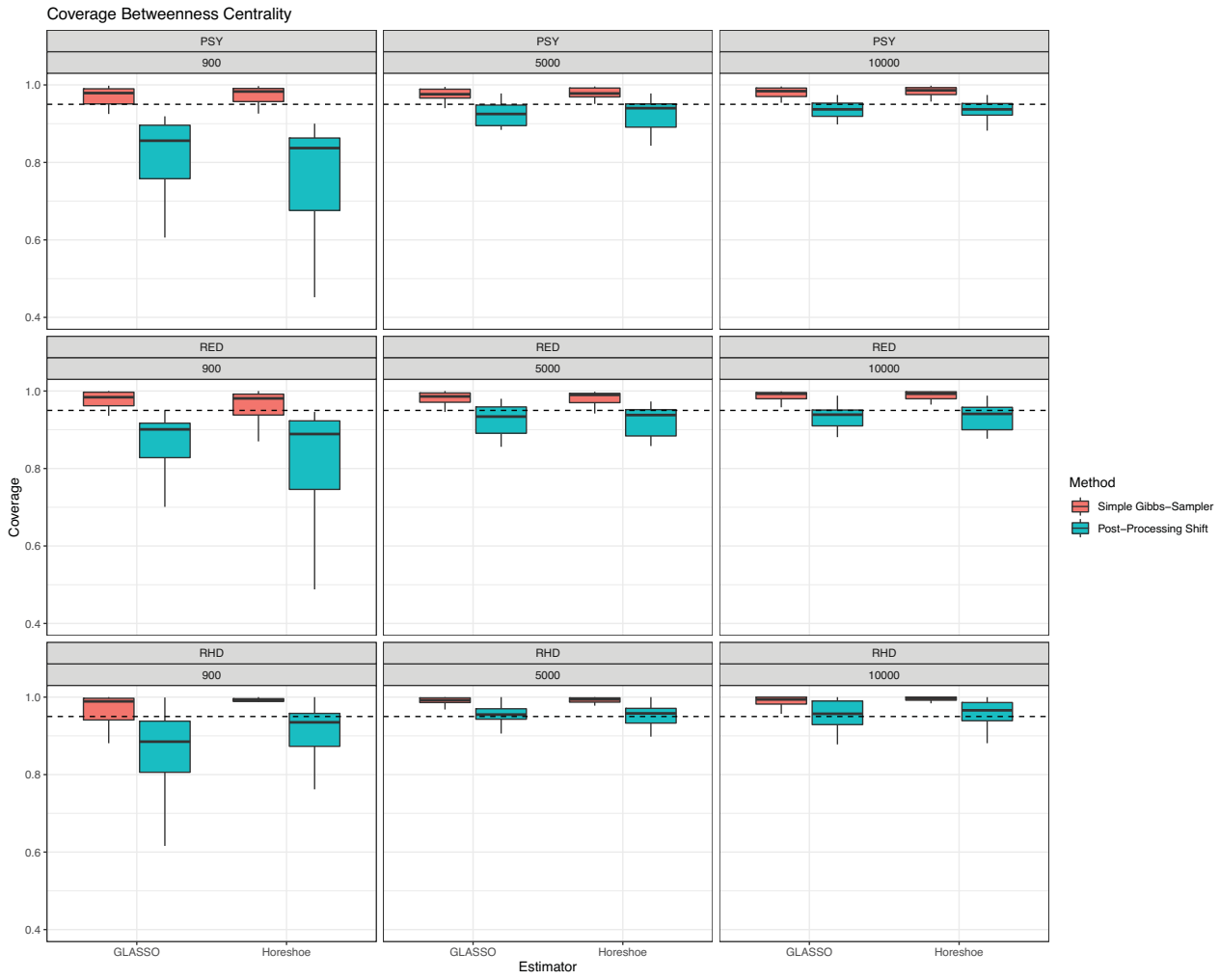


Figure 9. Coverage for betweenness centrality.

(due to it's sparser nature), which could pose a problem for the assumption of normality that this method is based on.

When using the graphical Horseshoe, closeness centrality is estimated with a little more bias than with the Bayesian GLASSO, however coverage is better for all methods except the Post-Processing Shift method (see Table 2, Tables A.4–A.6 of the online Appendix, and Figures 6a and 6b). When using the Graphical Horseshoe, coverage of the Post-Processing Shift method is only sufficient for $N > 900$ in the Psych and RED network, and never sufficient for the RHD network. The coverage for the estimation based method with correlated successive paths shows lower average and per node coverage than when the Bayesian GLASSO is used, but this method does have good average and per node coverage in the RHD network now for $N > 900$. The Simple Gibbs-Sampler estimation method again works a little better with the Horseshoe prior and now has good average and per node coverage for $N = 10{,}000$ in all three networks, and for $N > 900$ in the RED network.

Betweenness

Out of the 3 centrality estimates, betweenness centrality was estimated the least well by all methods. When using the Bayesian GLASSO, the Post-Processing Shift Estimation method again tends to have the lowest MSE (except for $N = 10{,}000$ in the RHD network) (see Tables A.1–A.3 of the online Appendix, and Figures 8a and 8b).

Unlike for strength and closeness centrality, coverage for betweenness centrality is not sufficient with any method. Also, coverage for the Post-Processing Shift estimation method is worse than that of the Simple Gibbs-Sampler (Table 1). This last method has better average coverage rates, and only has insufficient average coverage for $N > 900$ in the RHD network, where average coverage is too conservative. The per node coverage of the Simple Gibbs-Sampler is insufficient however for

all networks and sample sizes, although at $N > 900$ the method tends to be too conservative.

When using the Graphical Horseshoe, the MSE's for the betweenness centrality estimates again tend to be larger than when the Bayesian GLASSO is used (see Tables A.4–A.6 of the online Appendix, and Figures 8a and 8b). With this prior the MSE of the Post-Processing Shift method tend to be smaller than that of the Simple Gibbs-Sampler method for the denser Psych and RED networks (except when $N = 900$). For the RHD network, the Simple Gibbs-Sampler always has the lowest MSE. In terms of coverage the overall picture is similar to that seen with the Bayesian GLASSO (although coverage for Post-Processing Shift method appears somewhat lower, and that of the Simple-Gibbs Sampler method somewhat higher) (see Table 2). The average and per node coverage for the Post-Processing Shift method is too low, while the Simple Gibbs-Sampler method has good average coverage, with average coverage only being insufficient for $N > 900$ in the RHD network (where average coverage is too conservative). In addition, as was the case when the Bayesian GLASSO was used, the per node coverage of the Simple Gibbs-Sampler is insufficient for all networks and sample sizes, although at $N > 900$ the method tends to be too conservative.

Conclusion

Taken together, results of our simulation study shows that the Bayesian methods (particularly the Bayesian GLASSO) are strong alternatives for the frequentist GLASSO. The Bayesian GLASSO outperforms the frequentist GLASSO with respect to a) bias in edge weights, b) bias in the centrality measures, and c) the correlation between the estimated and true partial-correlations. The Bayesian Horseshoe also typically outperforms the frequentist GLASSO on these three measures (except at $N = 900$ where the frequentist GLASSO shows less bias in edge weights and higher correlations between estimated and true values).

In terms of sensitivity and specificity results are more mixed. The frequentist GLASSO has better sensitivity overall (although the Bayesian GLASSO is very close or slightly better in the RED and RHD networks for $N \geq 5,000$), while the Bayesian methods have better specificity. A choice between methods here therefore appears to come down to a preference for erring on the side of caution (deciding an edge is absent while it isn't) versus erring on the sides of discovery (deciding an edge is there while it isn't). However, the Bayesian GLASSO does perform better than the frequentist GLASSO (except at $N = 900$ in the Psych Network) on F1 which gives a harmonic mean of precision (the proportion of truly positive edges out of all the edges identified as positive by a method) and recall (the proportion of all true positive edges that gets identified as positive), while the Graphical Horseshoe performs better on this measure for $N \geq 900$.

In terms of (coverage for) the three different centrality measures, the Bayesian GLASSO outperform the Graphical Horseshoe, and shows good coverage for strength and closeness centrality at sample sizes of $N = 5,000$ or higher (in which case there are 16.66 observations for each partial correlation with 25 nodes). For smaller sample sizes the Bayesian GLASSO has insufficient coverage for strength centrality in the denser Psych- and RED networks, and insufficient coverage for closeness centrality in the less dense RHD network. This is likely the case because there is less shrinkage at $N > 900$. As mentioned, Bayesian regularization is a two-step approach; 1) estimates are pulled toward zero (but *not* set to zero) by the GLASSO or Horseshoe priors, and 2) estimates whose 95% Credibility Intervals contain zero are set to 0. At $N > 900$, the shrinkage in the first of these steps is less "severe" than at $N = 900$, likely leading to less distorted posteriors. For strength and closeness centrality the Post-Processing Shift estimation method provided the best performance out of all methods tested with the Bayesian GLASSO in both coverage and MSE. Performance with regard to betweenness centrality was worse than for the other two centrality measures, and coverage is insufficient for both the Bayesian GLASSO and Graphical Horseshoe at all N for all networks. For this centrality measure the Simple-Gibbs estimation appear to be the best choice. It has a larger MSE than the Post-Processing Shift method, but shows better coverage rates and tends to have intervals that are too wide for $N \geq 900$, implying that it will err on the side of being too conservative.

Finally, performance of the Bayesian GLASSO (and the Bayesian methods in general) appears to be better when networks are not to sparse, but this should not pose a problem in social sciences. With regard to network-structure, the methods worked well for both the structure of the bfi-data from the *psych-package*, and for a random network.

Limitations

In this study we evaluated the performance of different estimation methods by looking at their ability to

recover an underlying true- or generating model. This is only one way of evaluating estimation methods, and some argue not the most important one (Cudeck & Henly, 2003), as in practice there are no true models, and not being able to recover a generating model in a simulation study does not mean that a method does not have other merits (like simplicity or interpretability) or cannot capture relevant aspects of underlying processes in practical research (Cudeck & Henly, 2003; Hand & Vinciotti, 2003). Future research should also look at accuracy of model-based prediction and/or cross-validation as performance measures.

In addition, as is always the case with simulation studies, results only generalize within the scope of the settings and scenarios used in this study. That is, we only looked at a subset of possible data generating- and prior models. We did not consider an exhaustive set of different generating network structures because our focus was on the use of GGMs in the social-, and specifically, psychological sciences. We therefore wanted to compare methods on network structures that researchers in these field are likely to encounter (such as the bfi-data). Also, when determining the effect of network structure on the performance of the different methods, by also generating data based on networks with a different structure and/or density as that of the bfi-data, we wanted to keep these variations in structure within the limits that they might be encountered by, and therefore relevant to, applied social scientists (e.g., the size of the edge weights and the number of variables). Getting a more fine grained picture of the effect of different network characteristics (i.e., size of edge weights, general structure, sparsity, etcetera) on the performance of different estimation methods is important however. Future research should therefore look at more diverse sets of data, including but not limited to i) networks whose precision matrices includes edge weights that are (mostly) very close to zero, ii) networks much denser than the ones used in this study, or even iii) from networks who's characteristics are less common (or perhaps even unlikely) in psychological practice (to determine boundary performance). Regarding priors, we chose to focus on the Bayesian GLASSO and Horseshoe priors, again because of our focus on the use of GGMs in the social sciences. These two priors are currently quite popular in that field and therefore often used. However, many alternatives to these two choices exist. One interesting alternative is the already mentioned G-Wishart prior (Dawid & Lauritzen, 1993; Roverato, 2000), a continuous and discrete mixture prior that can set edges to 0 without the need for a heuristic like the one we used in this article. In addition, future research might also want to consider priors with Ridge-type penalization as in the Eigenvalue decomposition based Wishart prior introduced by Kuismin and Sillanpää (2016). Ridge penalization tends to shrink to 0 less strongly than LASSO regularization, and could therefore lead to better results with respect to sensitivity, albeit likely at the cost of specificity. Studying alternative priors will give valuable information about how they compare to each other and what analytic goals are best served by what prior. As suggested above, the GLASSO could be preferred when specificity is more important for example, while Ridge-type priors could be the better choice if sensitivity is more important. Lastly, future research might also want to study variations on the priors used in this study in more detail. For example, by setting different hyper-priors on the tuning parameters of the GLASSO and Horseshoe priors, or by using different heuristics for setting edges to 0, to further investigate the impact of choices in those aspects of the priors in different contexts.

In addition, we only looked at forms of regularization in which edges are set to exactly 0. As mentioned, Bayesian regularization does not set elements of the precision matrix to exactly 0 by default. It pulls estimates toward zero, but does not make them exactly zero. This less strict regularization might actually be a better option in some circumstances, depending on how realistic the assumption of sparsity is for the data (or population) at hand. Another, interesting difference between Bayesian regularization and frequentist regularization, is that, as Bayesian regularization is achieved by means of priors, the amount of regularization diminishes as the sample size goes up. This is actually really useful behavior. The more signal there is in the data, the less estimates are pulled to 0. Frequentist regularization also prunes less edges as data increases, as the EBIC will select denser graphs, but how the amount of regularization depends on the amount of data, and how this dependence is different than for the Bayesian methods is not very clear. Looking into Bayesian regularization (and how it compares to it's Frequentist cousin) in more detail, both from an applied and more theoretical point of view, is therefore needed.

Next to looking more closely into the differences between frequentist and Bayesian regularization mentioned above, future research could also look into non-regularized estimation of GGMs in more detail (Liang et al., 2015; Williams et al., 2019). Regularization makes inference more difficult as the

meaning of p-values becomes less clear. What is the exact null-hypothesis being tested, for example? And how can one account for the fact that the data is used not only for model estimation, but also for the process used to determine the penalty term (and therefore the sparsity of the final model)? As mentioned in the introduction, usually the presence of an edge is taken as proof for the presence of a relation between nodes, while the absence is taken as proof for conditional independence. This might seem to side-step the mentioned issues with p-values, but doesn't really. For every method (not just regularized methods) 1 – sensitivity gives the false negative rate, which indicates we cannot simply use the absence of an edge/connection as evidence the relationship is null. Similarly, 1 – specificity gives the false positive rate (and the frequentist GLASSO has been show to have a quite a few false positives, although this is less the case for the Bayesian GLASSO, see above), which indicates we cannot simply use the presence of an edge/connection as evidence the relationship exists. Non-regularized estimation could circumvent some of the inference problems mentioned above (although issues with post-selection inference apply to non-regularized estimation as well (Berk et al., 2013)), but it's performance (compared to regularized estimation) has not been extensively studied yet.

Finally, there is some debate on whether the different centrality indices should be used in the context of psychological networks (Bringmann et al., 2019). In their paper Bringmann et al. (2019), observe that centrality indices display wide confidence intervals (Bringmann et al., 2013), low stability in cross-sectional data (Epskamp et al., 2017), inconsistency (regarding which node is most central) across data sets (Bringmann et al., 2016; Forbes et al., 2017; however also see Borsboom et al., 2017), and are not always linked to external measure of interest (Rodebaugh et al., 2018). In addition, Bringmann et al. (2019) also discuss more conceptual difficulties with the actual meaning of the different centrality estimates in the context of psychological networks. Particularly betweenness and closeness centrality appear to be unsuitable as measures of node importance in psychological networks context. As mentioned in the introduction however, proper inference based on network models requires taking the accuracy of the estimates, including centrality measures, into account. Until now it wasn't possible to do this properly, which is why we looked into the Bayesian estimation of centrality indices in this paper. Proper uncertainty estimation for the centrality measures might make them more useful in practice, at least with regard to (consistently) ranking nodes based on centrality, and/or with regard to relating node centrality to external measures of interest or outcomes. For example, it could be that only when a node is clearly more central than another (based on it's 95% Credibility Interval and those of others), that it can be viewed as more important, but that so far identifying truly more central nodes was not possible due to the limitations in estimating the uncertainty of node centralities. In addition, when truly more central nodes can be identified from among all other nodes, centrality might also become more usefully related to external outcomes. The problem raised by Bringmann et al. (2019) on how the indices should be interpreted in the context of psychological networks, and on how they are unsuited to capture certain processes in a network won't be solved by proper uncertainty estimation however. In addition, our Bayesian methods also don't provide proper uncertainty estimates for betweenness centrality, so this measure will remain problematic regardless.

Article information

Conflict of interest disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was not supported.

Role of the funders/sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

References

Berk, R., Brown, L., Buja, A., Zhang, K., & Zhao, L. (2013). Valid post-selection inference. *The Annals of Statistics*, *41*(2), 802–837. https://doi.org/10.1214/12-AOS1077

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*(9), 1089–1108. https://doi.org/10.1002/jclp.20503

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121.

Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PLoS One*, *6*(11), e27407. https://doi.org/10.1371/journal.pone.0027407

Borsboom, D., Fried, E. I., Epskamp, S., Waldorp, L. J., van Borkulo, C. D., van der Maas, H. L. J., & Cramer, A. O. J. (2017). False alarm? A comprehensive reanalysis of "evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger (2017). *Journal of Abnormal Psychology*, *126*(7), 989–999. https://doi.org/10.1037/abn0000306

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144–152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., Wigman, J. T. W., & Snippe, E. (2019). What do centrality measures measure in psychological networks. *Journal of Abnormal Psychology*, *128*(8), 892–903. https://doi.org/10.1037/abn0000446

Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., Tuerlinckx, F., & Kuppens, P. (2016). Assessing temporal emotion dynamics using networks. *Assessment*, *23*(4), 425–435.

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS One*, *8*(4), e60188. https://doi.org/10.1371/journal.pone.0060188

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparsesignals. *Biometrika*, *97*(2), 465–480. https://doi.org/10.1093/biomet/asq017

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

Cramer, A. O. J., Waldorp, L., Maas, H. v. d., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, *33*(2–3), 137–150. https://doi.org/10.1017/S0140525X09991567

Cudeck, R., & Henly, S. (2003). A realistic perspective on pattern representation in growth data: Comment on Bauer and Curran. *Psychological Methods*, *8*(3), 378–383.

Dawid, A., & Lauritzen, S. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Annals of Statistics*, *21*, 1272–1317.

Dijkstra, E. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*(1), 269–271. https://doi.org/10.1007/BF01386390

Epskamp, S. (2020). *psychonetrics: Structural equation modeling and confirmatory network analysis [Computer software manual]* (R package version 0.8). https://CRAN.R-project.org/package=psychonetrics

Epskamp, S., Borsboom, D., & Fried, E. I. (2017). Estimating psychological networks and their accuracy: a tutorial paper. *Behavior Research Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 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*(4), 617–634. https://doi.org/10.1037/met0000167

Epskamp, S., Rhemtulla, M. T., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*(4), 904–927. https://doi.org/10.1007/s11336-017-9557-x

Forbes, M., Wright, A., Markon, K., & Krueger, R. (2017). Further evidence that psychopathology networks have limited replicability and utility: Response to Borsboom et al. (2017) and Steinley et al. (2017). *Journal of Abnormal Psychology*, *126*(7), 1011–1016.

Friedman, J. H., Hastie, T., & Tibshirani, R. (2008). Sparse inversecovariance estimation with the graphical lasso. *Biostatistics (Oxford, England)*, *9*(3), 432–441.

Hand, D. J., & Vinciotti, V. (2003). Local versus global models for classification problems. *The American Statistician*, *57*(2), 124–131. https://doi.org/10.1198/0003130031423

Kuismin, M., & Sillanpää, M. (2016). Use of wishart prior and simple extensions for sparse precision matrix estimation. *PLoS One*, *11*(2), e0148171. https://doi.org/10.1371/journal.pone.0148171

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

Li, Y., Craig, B. A., & Bhadra, A. (2019). The graphical horseshoe estimator for inverse covariance matrices. *Journal of Computational and Graphical Statistics*, *28*(3), 747–757. https://doi.org/10.1080/10618600.2019.1575744

Liang, F., Song, Q., & Qiu, P. (2015). An equivalent measure of partial correlation coecients for high-dimensional gaussian graphical models. *Journal of the American Statistical Association*, *110*(511), 1248–1265. https://doi.org/10.1080/01621459.2015.1012391

Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs andvariable selection with the lasso. *The Annals of Statistics*, *34*(3), 1436–1462. https://doi.org/10.1214/009053606000000281

Mood, A. M., Graybill, F. A., & Boes, D. C. (1985). *Introduction to the theory of statistics* (3rd ed.). McGraw-Hill.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251. [Database] https://doi.org/10.1016/j.socnet.2010.03.006

Revelle, W. (2019). *psych: Procedures for psychological, psychometric, and personality research [Computer software manual]* (R package version 1.9.12). https://CRAN.R-project.org/package=psych

Rodebaugh, T. L., Tonge, N. A., Piccirillo, M. L., Fried, E., Horenstein, A., Morrison, A. S., Goldin, P., Gross, J. J., Lim, M. H., Fernandez, K. C., Blanco, C., Schneier, F. R., Bogdan, R., Thompson, R. J., & Heimberg, R. G. (2018). Does centrality in a cross-sectional network suggest intervention targets for social anxiety disorder. *Journal of Consulting and Clinical Psychology*, *86*(10), 831–844. https://doi.org/10.1037/ccp0000336

Rothman, A. J., Levina, E., & Zhu, J. (2010). Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics*, *19*(4), 947–962. https://doi.org/10.1198/jcgs.2010.09188

Roverato, A. (2000). Cholesky decomposition of a hyper-inverse wishart matrix. *Biometrika*, *87*(1), 99–112. https://doi.org/10.1093/biomet/87.1.99

Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: a network perspective on psychological phenomena. *New Ideas in Psychology*, *31*(1), 43–53. https://doi.org/10.1016/j.newideapsych.2011.02.007

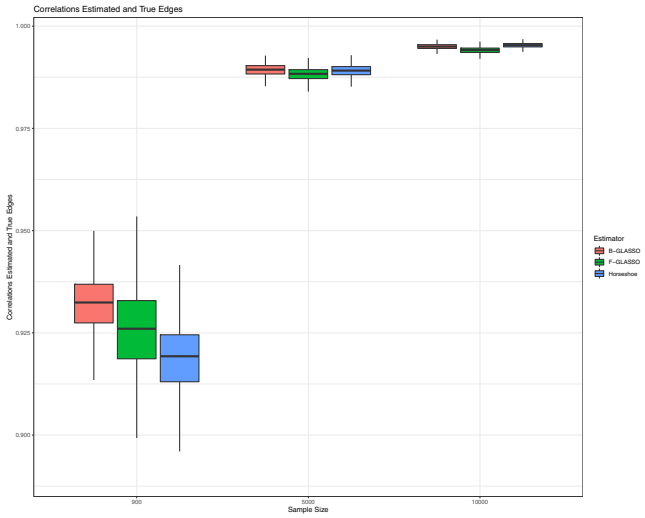
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Wang, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis*, *7*(4). https://doi.org/10.1214/12-BA729

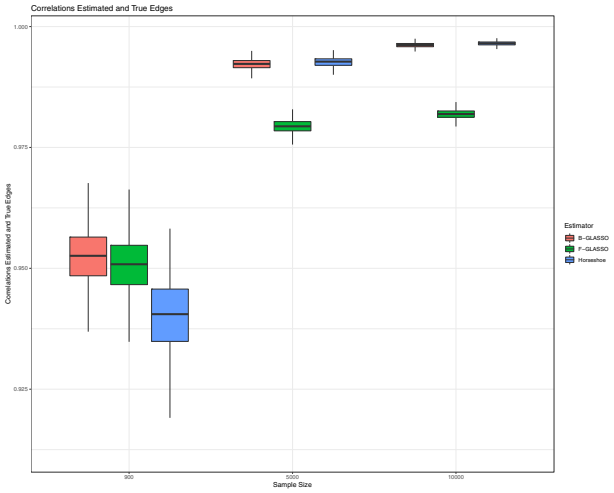
Williams, D., Rhemtulla, M., Wysocki, A., & Rast, P. (2019). On nonregularized estimation of psychological networks. *Multivariate Behavioral Research*, *54*(5), 719–750.

Yuan, M., & Lin, Y. (2007). Model selection and estimationin the gaussian graphical model. *Biometrika*, *94*(1), 19–35. https://doi.org/10.1093/biomet/asm018

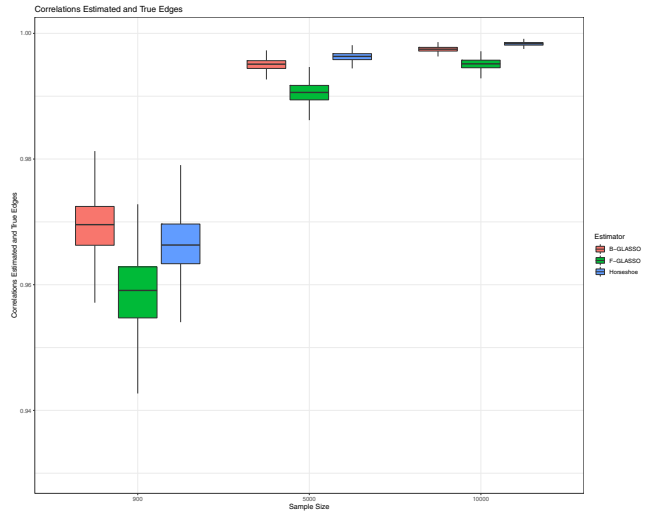
Appendix A



a Correlation Between Estimated and True Edges for Psych Network



b Correlation Between Estimated and True Edges for Random Network



c Correlation Between Estimated and True Edges for Random Network (Half Density)

Figure A1. Correlations Between Estimated and True Edge Weights.

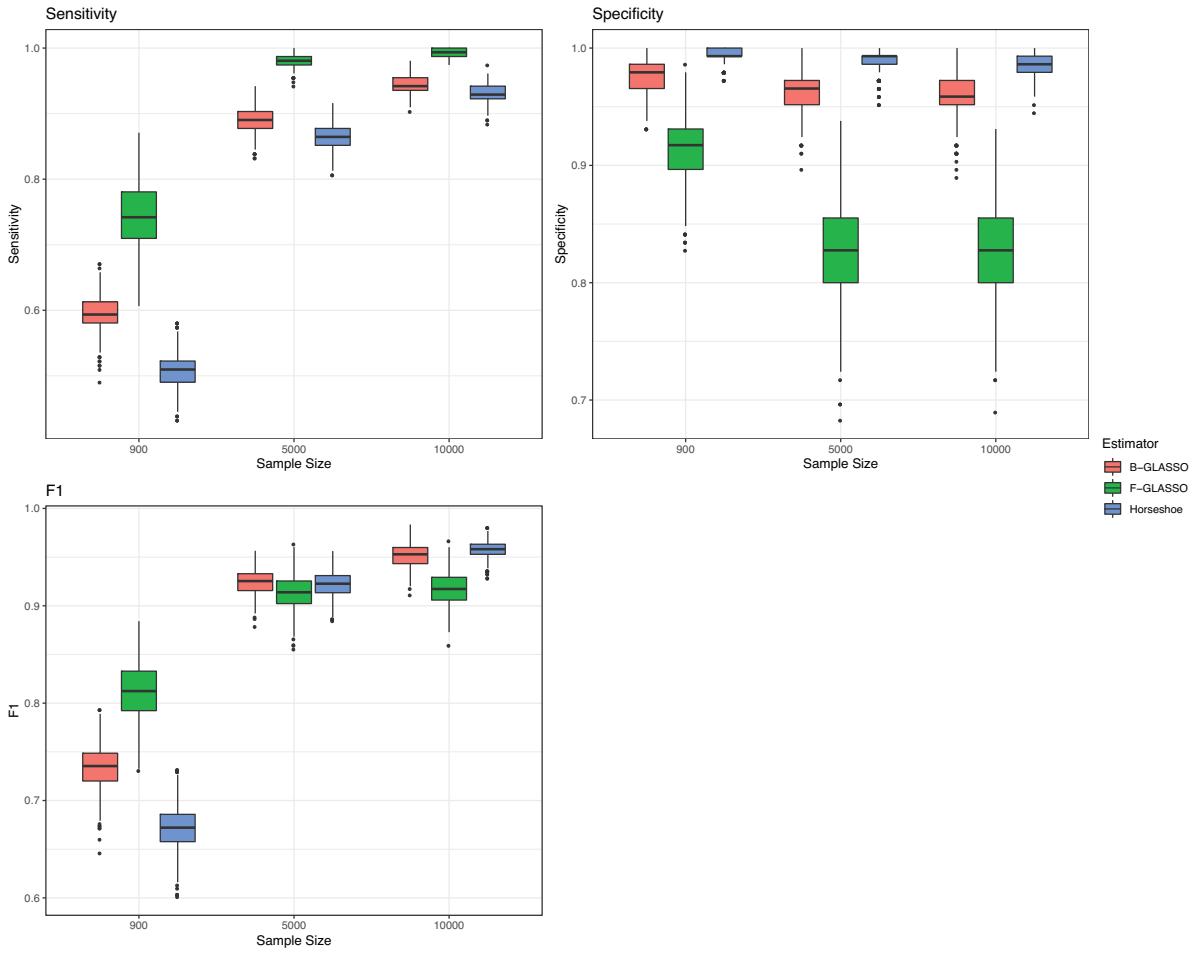


Figure A2. Sensitivity, specificity, and F1 for the Psych Network.

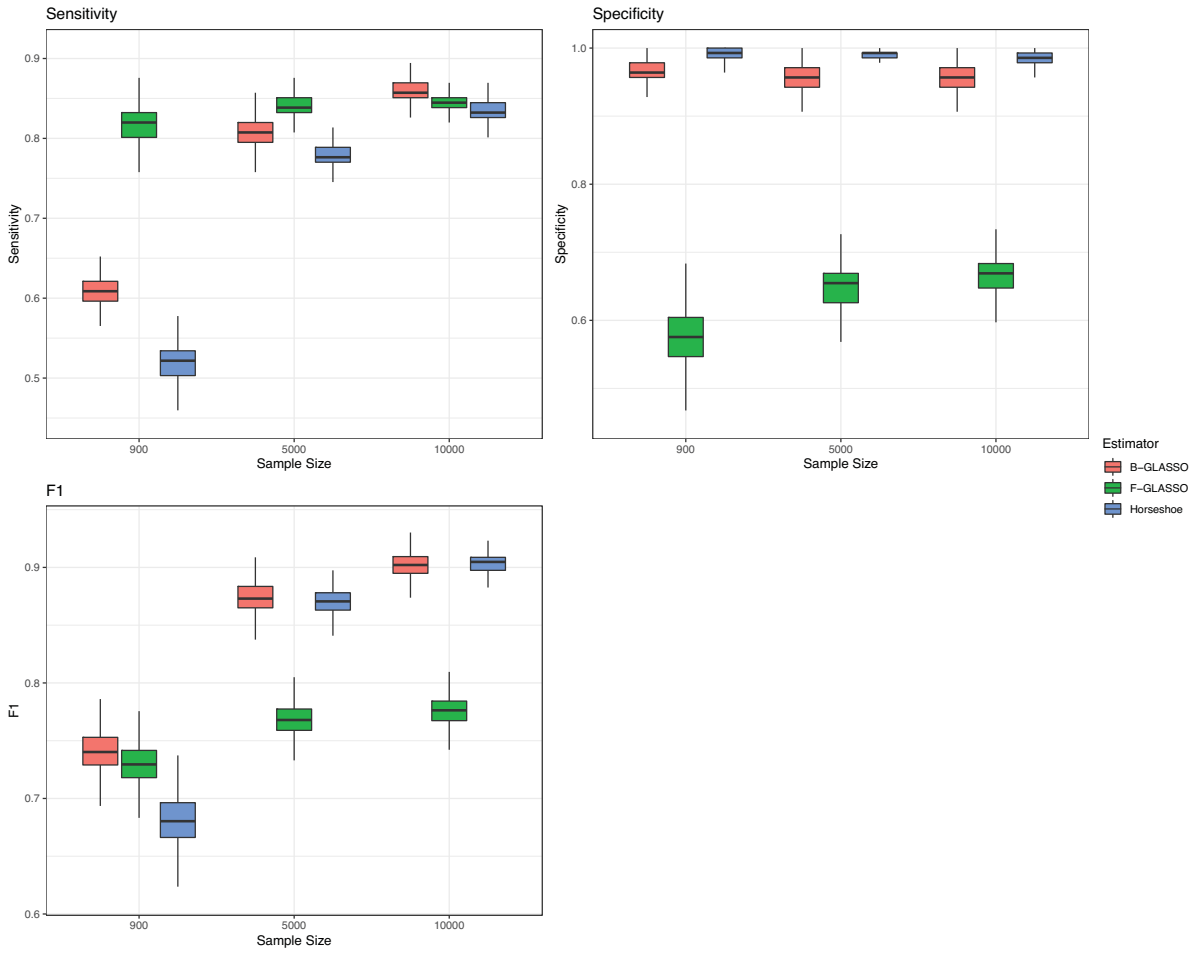


Figure A3. Sensitivity, specificity, and F1 for the random network.

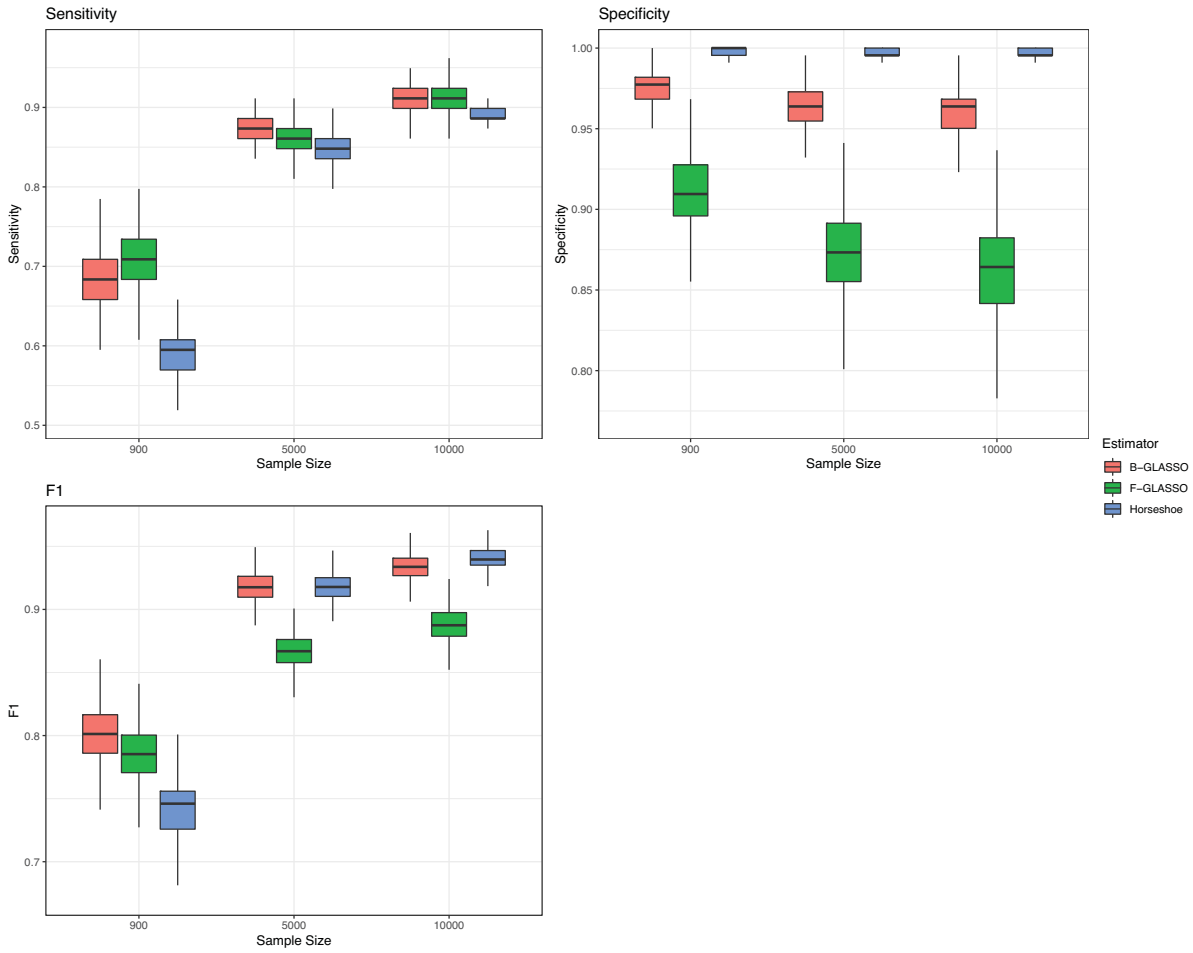


Figure A4. Sensitivity, specificity, and F1 for the random network (half density).

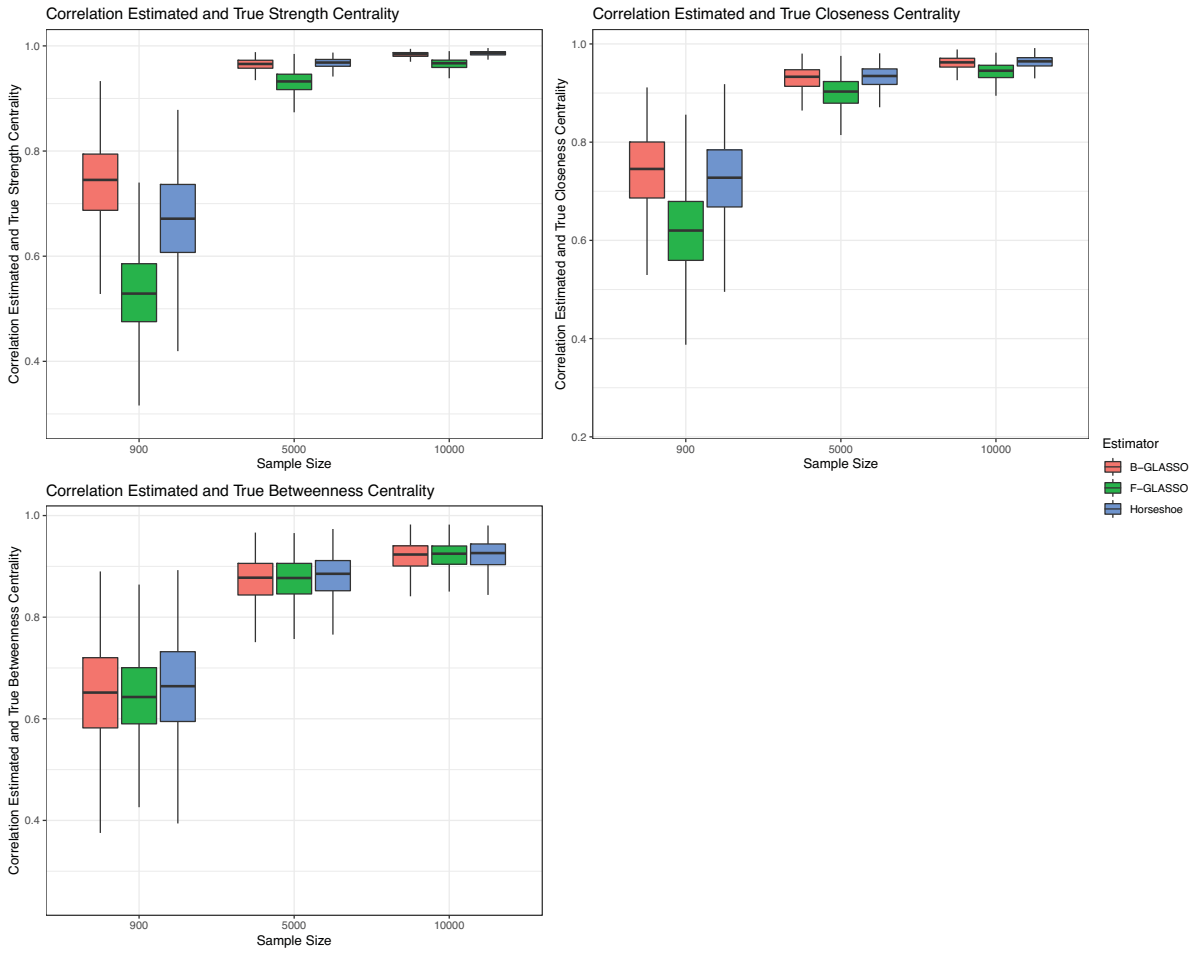


Figure A5. Correlation between the estimated and true centrality values for the Psych Network.

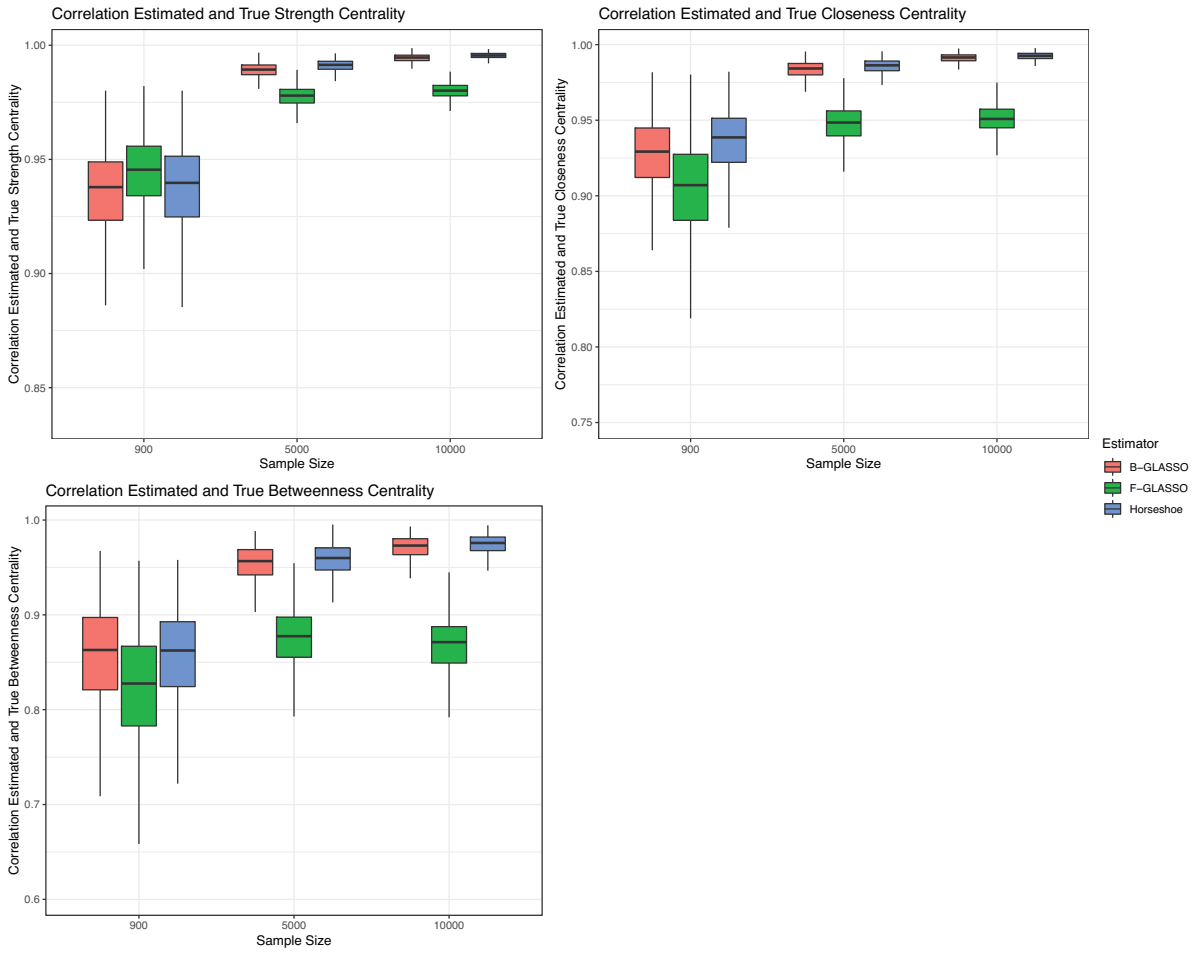


Figure A6. Correlation between the estimated and true centrality values for the random network.

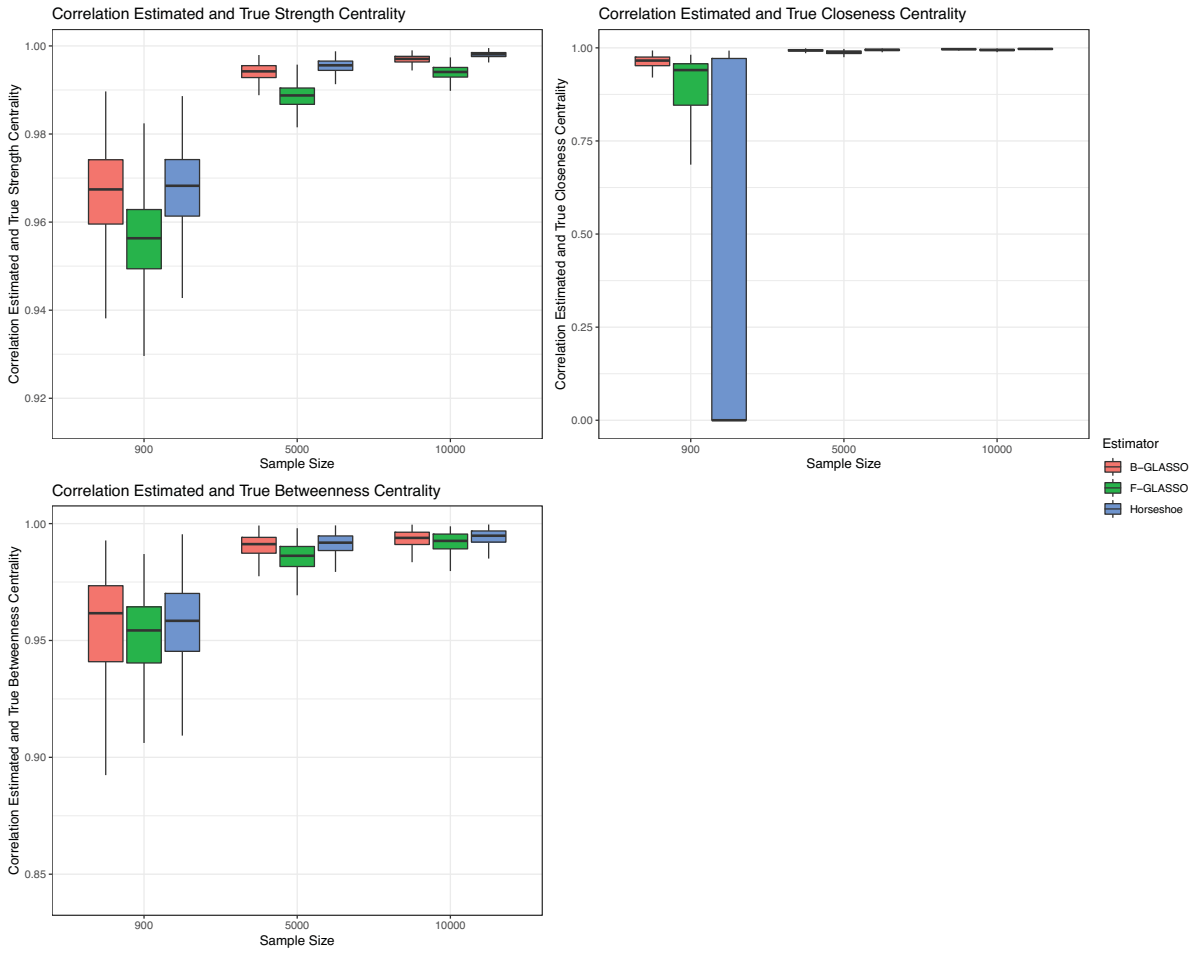


Figure A7. Correlation between the estimated and true centrality values for the random network (half density).