Author Manuscript

HHS Public Access

Author manuscript

Comput Stat Data Anal. Author manuscript; available in PMC 2015 December 01.

Published in final edited form as:

Comput Stat Data Anal. 2015 September ; 89: 126–133. doi:10.1016/j.csda.2015.02.012.

Estimating a graphical intra-class correlation coefficient (GICC) using multivariate probit-linear mixed models

Chen Yue[a], **Shaojie Chen**[a], **Haris I. Sair**[b], **Raag Airan**[b], and **Brian S. Caffo**[a,*]

[a]Department of Biostatistics, Johns Hopkins University, Baltimore, MD, USA, 21205

[b]Department of Radiology and Radiological Science, Johns Hopkins University, Baltimore, MD 21205

Abstract

Data reproducibility is a critical issue in all scientific experiments. In this manuscript, the problem of quantifying the reproducibility of graphical measurements is considered. The image intra-class correlation coefficient (I2C2) is generalized and the graphical intra-class correlation coefficient (GICC) is proposed for such purpose. The concept for GICC is based on multivariate probit-linear mixed effect models. A Markov Chain Monte Carlo EM (mcm-cEM) algorithm is used for estimating the GICC. Simulation results with varied settings are demonstrated and our method is applied to the KIRBY21 test-retest dataset.

Keywords

graphical intra class correlation coefficient; multivariate probit-linear mixed model; MCMCEM

1. Introduction

A crucial question in any statistical analysis is: how reliable is the data? Experimental replication for the purpose of measuring the reliability of measurements is the most common method for establishing reproducibility. In this paper, we consider repeated measurement of graphs and propose the concept of the graphical intra-class correlation coefficient for measuring their reliability.

The Intra-class correlation coefficient (ICC) has been proposed [1] and used to evaluate the reliability of measurements in a variety applications [2], [3]. ANOVA mixed-effect models have been proposed [4] as a framework for estimating the ICC. Suppose y_{ij} denotes the j^{th} measurement of subject i, x_i denotes the subject specific random effect and u_{ij} indicates the measurement error. The one-way ANOVA model is:

$$y_{ij}=\mu+x_i+u_{ij} \quad x_i\sim N(0,\sigma_x^2), u_{ij}\sim N(0,\sigma_u^2), i.i.d. \qquad (1)$$

The ICC is then defined as:

*Corresponding author 615 N. Wolfe Street, E3610, Baltimore, MD, USA, 21205 (Brian S. CAFFO).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$ICC = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}. \quad (2)$$

In (1) and (2), the total variability of the data is decomposed into subject-specific variability and measurement error; ICC represents the proportion of variability that is due to heterogeneity in subjects. In recent research, the ICC has been generalized to multivariate cases. The work in [5] proposed a model analogous to (1) in functional data using multilevel functional principal component analysis (MFPCA) and an image intra-class correlation coefficient (I2C2) was subsequently proposed in [6] to calculate ICC for image data.

Graphical data are becoming increasingly popular in scientific research. Notably, graphs are used in describing brain networks in neuroimaging. In such research, binary graphs are often obtained from functional magnetic resonance image (fMRI) [7], [8], [9], [10], [11]. The increasing number of graphical datasets motivates us to evaluate the reliability of binary graphs.

Figure. 1 illustrates idealized graphical measurements for three different subjects. Here each subject is measured three times. The left panel shows a case where graphical measurements resemble each other within one subject. The ICC, consequentially, should be higher. The right panel, on the other hand, demonstrates the opposite situation, where the repeated measurements within one subject show poor consistency. In such case, the ICC should be relatively lower. In this manuscript, we propose the concept of the graphical ICC (GICC) to quantify the similarity between repeated measurements of binary graphs. In Figure. 1, each binary graph is represented by a 0 – 1 vector. For example, the first graph of subject 1 is represented by $(1, 1, 0, 1, 0, 0)^T$[1]. Thus our goal is to define an ICC for multivariate binary data.

Many authors have discussed the ICC for single variate binary data. [12] proposed a moment based estimator. Probit linear mixed-effect models were used by [13] and [14] to estimate a confidence interval for binary data ICC.

There is also work discussing the similarity between graphs. The work in [15] and [16] discussed the similarity between nodes and edges in graphs. One main purpose of these papers was to find assembled subgraphs between two graphs. Instead of having a fixed node-to-node or edge-to-edge match, they found the match between two graphs based on edge/node similarity score.

Our objective, on the other hand, is to estimate the ICC to evaluate the reliability of replicated measurement of binary graphs. In section 2, a multivariate probit linear mixed model is proposed. A Monte Carlo expectation maximization (MCEM) algorithm will be discussed in section 3. Simulation results with various settings will be shown in section 4 and the results of our method being implemented on binary brain connectivity maps are in section 5. We will summarize the paper in section 6.

[1] Each element of the vector is an indicator of the existence of an edge, the order of the six elements is ① – ②, ① – ③, ① – ④, ② – ③, ② – ④, ③ – ④.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2. Model

Suppose $\{o_{ij}(d) : i = 1, \ldots, I; j = 1 \ldots, J_i; d = 1, \ldots, D, \}$ are binary observations representing repeated graph measurements for multiple subjects. Here, I is the total number of subjects, J_i is the number of visits for the i^{th} subject and D is the number of possible edges for all graphs. Usually, we have $D=\frac{N(N-1)}{2}$ where N is the number of nodes. In Figure. 1, for example, we have $I = 3$, $J_i = 3$, $N = 4$, $D = 6$. The multivariate probit-linear mixed model is as follows:

$$\Phi^{-1}(P(o_{ij}(d)|x_i(d)))=\mu(d)+x_i(d),\ \mathbf{x}_i\sim\mathbf{N}(\mathbf{0},\mathbf{\Sigma}_x), \quad (3)$$

where $\mathbf{x}_i = (x_i(1), \ldots, x_i(D))^T$. The GICC, is then defined as:

$$GICC=\frac{tr(\mathbf{\Sigma}_x)}{tr(\mathbf{\Sigma}_x)+D}. \quad (4)$$

For the purpose of estimation, the model can also be viewed as a threshold model that dichotomizes the observations from a latent Gaussian distribution. In other words:

$$o_{ij}(d)=\mathbf{I}_{(y_{ij}(d)>0)},\ y_{ij}(d)=\mu(d)+x_i(d)+u_{ij}(d),\ \mathbf{x}_i\sim\mathbf{N}(\mathbf{0},\mathbf{\Sigma}_x), i.i.d.,\ \mathbf{u}_{ij}\sim\mathbf{N}(\mathbf{0},\mathbf{I}), i.i.d., \quad (5)$$

where $\mathbf{x}_i = (x_i(1), \ldots, x_i(d))^T$ and $\mathbf{u}_{ij} = (u_{ij}(1), \ldots, u_{ij}(d))^T$. The equivalency of these two models can be easily shown by the following calculation:

$$P(o_{ij}(d)=1|x_i(d))=P(y_{ij}(d)>0|x_i(d))=P\left(u_{ij}(d)>-(\mu(d)+x_i(d))\middle|x_i(d)\right)=1-\Phi(-\mu(d)+x_i(d)))=\Phi(\mu(d)+x_i(d)).$$

Formula 4 is a direct generalization from the univariate ICC Formula 2. $GICC = 0$ indicates that $tr(\Sigma_x) = 0$, which means that the between group variance is zero for all dimensions, $d = 1, \ldots, D$. $GICC \approx 1$ indicates that $tr(\Sigma_x) \gg D$, implying that the variation between subjects is much larger than the variation within subject. When $GICC = 0.5$, $tr(\Sigma_x) = D$, implying that the overall between subject variation is equal to the within subject variation.

The advantage of using the *trace* is that: (1) it is an overall statistic instead of a edge specific statistic, which provides a global measurement to quantify graphical reproducibility; (2) compared to other numerical methods (e.g. $\max(diag(\Sigma))$, $\sum(\sigma_{ij})$), the trace is invariant to orthogonal transformations, which is critical in measuring the variability of vectors.

Furthermore, using trace for measuring the global reproducibility was also proposed for the image ICC (I2C2) (see [6]) in and the functional version of ICC (see [5]).

3. The Monte Carlo EM Algorithm

MCEM algorithms have been used in probit-linear mixed models with single variate outcomes [17]. Here MCEM is generalized to the multivariate case. In model 5, the parameters of interest are μ and Σ_x. In the procedure of estimation, we treat **o** as observed data and [**y**, **x**] as the full data.

3.1. M-step

Given the full data **y** and **x**, the MLE for both parameters yields an explicit form:

$$\hat{\mu}=\frac{1}{\sum_i J_i}\sum_i\sum_j(\mathbf{y}_{ij}-\mathbf{x}_i),\ \hat{\mathbf{\Sigma}}_x=\frac{1}{I}\sum_i\mathbf{x}_i\mathbf{x}_i^T. \quad (6)$$

Unlike [18], the estimate of μ does not involve Σ_x, since **x** is also treated as part of the complete data. So $\hat{\mu}$ is obtained based on both **x** and **y**, rather than only on **y**.

Substituting **y**, **x** and $\mathbf{x}\mathbf{x}^T$ with $E[\mathbf{y}|\mathbf{o}]$, $E[\mathbf{x}|\mathbf{o}]$ and $E[\mathbf{x}\mathbf{x}^T|\mathbf{o}]$ respectively on the right side of 6, we obtain the M-step.

3.2. E-step

Based on 6, it is necessary to calculate $E(\mathbf{y}_{ij}|\mathbf{o})$, $E(\mathbf{x}_i|\mathbf{o})$ and $E(\mathbf{x}_i\mathbf{x}_i^T|\mathbf{o})$. Note that

$$E[\mathbf{x}_i|\mathbf{o}]=E[E[\mathbf{x}_i|\mathbf{y}]|\mathbf{o}],\ E[\mathbf{x}_i\mathbf{x}_i^T|\mathbf{o}]=E[E[\mathbf{x}_i\mathbf{x}_i^T|\mathbf{y}]|\mathbf{o}]. \quad (7)$$

The inner expectation can be obtained by using the joint distribution of $\{\mathbf{x}_i, \mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i}\}$. Noticing the following fact:

$$\begin{aligned}
&[\mathbf{x}_i,\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}]\\
&\quad=\prod_j[\mathbf{y}_{ij}|\mathbf{x}_i]\\
&\quad\times[\mathbf{x}_i]\propto\exp\{\\
&\quad-\frac{1}{2}\left[\sum_j\{(\mathbf{y}_{ij}-\mu-\mathbf{x}_i)^T(\mathbf{y}_{ij}\right.\\
&\quad\left.-\mu-\mathbf{x}_i)\}+\mathbf{x}_i^T\mathbf{\Sigma}_x^{-1}\mathbf{x}_i\right]\propto\exp\left\{-\frac{1}{2}\left[\mathbf{x}_i^T(J_i\mathbf{I}\right.\right.\\
&\quad+\mathbf{\Sigma}_x^{-1})\mathbf{x}_i\\
&\quad\left.\left.-2[\sum_j(\mathbf{y}_{ij}-\mu)]^T\mathbf{x}_i\right]\right\},
\end{aligned} \quad (8)$$

it can be derived that:

$$\mathbf{x}_i|\mathbf{y}_{i1},\ldots\mathbf{y}_{iJ_i}\sim\mathbf{N}\left((J_i\mathbf{I}+\mathbf{\Sigma}_x^{-1})^{-1}(\mathbf{y}_{i.}-J_i\mu),(J_i\mathbf{I}+\mathbf{\Sigma}_x^{-1})^{-1}\right), \quad (9)$$

where $\mathbf{y}_{i.}=\sum_j\mathbf{y}_{ij}$. Thus we have

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$E(\mathbf{x}_i|\mathbf{O})=E[E[\mathbf{x}_i|\mathbf{y}]|\mathbf{o}]=E\left[(J_i\mathbf{I}+\boldsymbol{\Sigma}_x^{-1})^{-1}(\mathbf{y}_{i.}-J_i\mu)|\mathbf{o}\right]=(J_i\mathbf{I}+\boldsymbol{\Sigma}_x^{-1})^{-1}(E[\mathbf{y}_{i.}|O]-J_i\mu),$$

$$\begin{aligned}E[\mathbf{x}_i\mathbf{x}_i^T|\mathbf{o}]&=E[E[\mathbf{x}_i\mathbf{x}_i^T|\mathbf{y}]|\mathbf{o}]\\&=E\left[(J_i\mathbf{I}+\boldsymbol{\Sigma}_x^{-1})^{-1}(\mathbf{y}_{i.}-J_i\mu)(\mathbf{y}_{i.}-J_i\mu)^T(J_i\mathbf{I}+\boldsymbol{\Sigma}_x^{-1})^{-1}+(J_i\mathbf{I}+\boldsymbol{\Sigma}_x^{-1})^{-1}|\mathbf{o}\right]\\&=(J_i\mathbf{I}+\boldsymbol{\Sigma}_x^{-1})^{-1}E[(\mathbf{y}_{i.}-J_i\mu)(\mathbf{y}_{i.}-J_i\mu)^T|\mathbf{o}](J_i\mathbf{I}+\boldsymbol{\Sigma}_x^{-1})^{-1}+(J_i\mathbf{I}+\boldsymbol{\Sigma}_x^{-1})^{-1}.\end{aligned}\quad(10)$$

However, the term $E[\mathbf{y}_{i.}|\mathbf{o}]$ and $E[\mathbf{y}_{i.}^T\mathbf{y}_{i.}|\mathbf{o}]$ does not have an explicit form. Here we use a Gibbs sampler to approximate the conditional expectation. Notice that, given $\mathbf{o}$, the distribution of $\mathbf{y}$ is multivariate truncated normal. The Gibbs sampler for such a distribution has been discussed in [19], [20], [21]. In the Gibbs sampling cycles, we choose the burn in period to be the first $T = 200$ and treat the following $B = 500$ elements as limiting realizations from the conditional distribution of $\mathbf{y}|\mathbf{o}$. Then an empirical conditional expectation is calculated as follows:

$$\hat{E}[\mathbf{y}_{ij}|\mathbf{o}]=\frac{1}{B}\sum_{b=T+1}^{T+B}\mathbf{y}_{ij}^{(b)},\quad(11)$$

$$\hat{E}[\mathbf{y}_{i.}\mathbf{y}_{i.}^T|\mathbf{o}]=\frac{1}{B}\sum_{b=T+1}^{T+B}\mathbf{y}_{i.}^{(b)}\mathbf{y}_{i.}^{(b)T}.$$

3.3. Observed information matrix for μ

Though we are not specifically interested in estimating μ for the graphical ICC, the estimate of μ with its standard error remains of potential interests, especially for modeling multivariate binary data using probit-linear mixed model. [22] expressed the observed information matrix in EM algorithm using the first and second derivative of the full likelihood.

Assume the observed log-likelihood is $l_o(\mathbf{o}, \theta)$ where $\theta = (\mu, \Sigma_x)$ and the full log-likelihood is $l_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y}, \theta)$, following [22], we have:

$$I_{\mathbf{o}}(\theta)=E_\theta\left[-\frac{\partial^2 l_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y},\theta)}{\partial\mu\partial\mu^T}|\mathbf{o}\right]-E_\theta\left[\left(\frac{\partial l_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y},\theta)}{\partial\mu}\right)\left(\frac{\partial l_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y},\theta)}{\partial\mu}\right)^T\middle|\mathbf{o}\right]+\left(\frac{\partial l_o(\mathbf{o},\theta)}{\partial\mu}\right)\left(\frac{\partial l_{\mathbf{o}}(\mathbf{o},\theta)}{\partial\mu}\right)^T.\quad(12)$$

Let $I_{\mathbf{o}} = I_{\mathbf{o}}(\hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimator. Then we have:

$$I_{\mathbf{o}}(\hat{\theta})=E_{\hat{\theta}}\left[-\frac{\partial^2 l_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y},\theta)}{\partial\mu\partial\mu^T}|\mathbf{o}\right]-E_{\hat{\theta}}\left[\left(\frac{\partial l_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y},\theta)}{\partial\mu}\right)\left(\frac{\partial l_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y},\theta)}{\partial\mu}\right)^T\middle|\mathbf{o}\right].\quad(13)$$

Following the same path as in the E-step, we can use Gibbs sampler and empirical averages to approximate the conditional expectation.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

4. Simulation

4.1. Estimates

We set number of subjects at I = 100, 200 and each subject receives J = 2, 4 repeated measurements. The number of nodes is set to be N = 5 so that the number of possible undirected edges is D = 10. The true μ is set to be 0.5 for all elements and

$$\Sigma_x[i,j]=r\rho^{|i-j|}, \text{where} \rho=0.8.$$

The underlying true graphical ICC using definition 4 is controlled by r. We set r = 2, 4 in each setting so that the corresponding ICC's are $\frac{rD}{rD+D}=2/3$ and 4/5 respectively. A total of 500 simulations were run in each simulation group.

In Table. 1, the average estimated GICC for r = 2 groups are 0.702, 0.672, 0.683 for $I_{100}J_2$, $I_{100}J_4$ and $I_{200}J_4$ group respectively, comparing to an under-lying truth 2/3 ≈ 0.667. As number of individuals increases, or as the number of repeated measurements increases, both the bias and the standard deviation of the estimated GICC reduces. When r = 4, the average estimated graphical ICCs are 0.817, 0.800 and 0.806, respectively. The MLE of GICC in each case has a positive bias, which is reduced as either I or J_i increases.

4.2. Robustness

In the proposed multivariate probit-linear mixed model, the assumption was made that the underlying within group error term is independent across edges. In other words, in model 5, the off-diagonal elements of $var(\mathbf{u}_{ij})$ is set to be zero. In real applications, this may not be the case. Therefore, we conducted simulations using the above settings when I = 100, J = 2 and r = 2. However, instead of setting Σ = I, we set $\sigma_{ij} = \rho^{|i-j|}$, where ρ = 0.1 and 0.5. The resulting GICC based on 100 simulations have means 0.699(0.025) and 0.722(0.026). The results highlight that the resulting GICC is still close to the underlying truth when there are correlations between edges in the error term.

4.3. Comparison With Other Benchmarks

We compared our proposed method with other available methods using our first simulation case (I = 100; J = 2; r = 2). The work in [23], [24] and [25] used ICC derived from one-way ANOVA, [13] proposed ICC for binary data using a single variate probit model. It should be pointed out that, all the other methods are based on the single variate ICC, so the comparison is limited. We compare our method with (1) average ICC(1)'s for all edges based on one-way ANOVA, (2) ICC(1) for the mean of the binary vector and (3) average edge-wise ICC based on single variate probit model. The results are shown in Table 2.

First of all, the first two ICCs are all derived from the one-way ANOVA model, thus all of the binary data are considered to be continuous. The average edge-wise ICC is only 0.457, but ICC for the mean vector is 0.812. None of them are close to the truth. The edge-wise ICC treats all binary data as pure jumps instead of treating them as having underlying continuous data. The ICC for the mean of vector can only provide the ICC on the average

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

statistic, but can not align the data with the same edge. The average edge-wise binary ICC based on single variate probit model also yields lower ICC. It also shows the difference between an unstructured covariance matrix for x and a diagonal matrix.

4.4. Running Time

To evaluate the running time of the MCEM algorithm, we conducted simulations with varied settings in which the number of subjects = 50, 100, the number of replicates = 2, 3, 4 and the number of edges = 10, 20 and 30. To be consistent with all settings, we terminated the algorithm after 30 iterations. All simulations were conducted using a 2.4GHz core on PowerEdge C6145 AMD Processor-based 2U Rack Server.

Table 3 shows that the number of subjects has a strict linear relationship with running time. Running time increases nonlinearly with either the number of replicates or the number of edges in the graph. With under 30-edge graphs and 4 replicates, the running time is less than twenty minutes. It is clear that if the number of replicates is 4, the running time grows faster than a quadratic function of the number of edges. Therefore, the current algorithm requires a relative small number of edges for each graph. We will further discuss it in Section 6.

5. Application

Resting-state fMRI scans consisted of a test-retest dataset previously acquired at the FM Kirby Research Center at the Kennedy Krieger Institute, Johns Hopkins University [26] are used to highlight the method. Twenty one healthy volunteers with no history of neurological disease each underwent two separate resting state fMRI sessions on the same scanner. A 3T MR scanner was used (Achieva, Philips Healthcare, Best, The Netherlands) utilizing a body coil with a 2D echoplanar (EPI) sequence and eight channel phased array SENSitivity Encoding (SENSE; factor of 2) with the following parameters: TR 2s, 3mm × 3mm in plane resolution, slice gap 1mm, for total imaging time of 7 minutes and 14 seconds. One subject was excluded due to technical issues at acquisition.

ICA (Independent Component Analysis) was performed using MEDOLIC (Multivariate Exploratory Linear Optimized Decomposition into Independent Components) version 3.10 in FSL (FMRIB Software Library, FMRIB, Oxford, UK). Preprocessing included removal of low-frequency drift with a highpass filter cutoff of 250s, realignment of the fMRI time series using MCFLIRT, slice timing correction, brain extraction using BET, and spatial smoothing with FWHM of 6mm. Images were registered to MNI standard space with resampling resolution of 2mm. ICA was performed using multi-session temporal concatenation with automatic dimensionality estimation and time-course variance normalization implemented in MELODIC. 43 components were identified by MELODIC.

Relevant ICA components corresponding to known large scale brain networks were identified by a board certified neuroradiologist with experience in resting state fMRI. Seven total components were selected (default mode network, dorsal attention network, motor network, visual network, salience network, and two lateralized executive control networks), and the 7 by 7 correlation matrix was calculated (see raw data in Figure 5). Different thresholds were used to dichotomize the raw graphs into binary ones, where the thresholds

Author Manuscript

were chosen from 0.1 to 0.8 using grid 0.01 (see Figure 5). The GICC algorithm was then implemented on these binary graphs.

The GICC was then calculated for each threshold (see Figure 3). The GICC remains above 0.6 when the threshold is between 0.1 and 0.6. For threshold outside of this band the GICC decreases dramatically. When the threshold is around 0.8, GICC fluctuates more significantly and the value eventually drops to 0.1. Thus the GICC shows high reproducibility of the raw data if a reasonable threshold is employed (from 0.1 to 0.6). When the threshold is too high, only few raw values will be dichotomize to 1, such that poor reproducibility is obtained. For practical subsequent applications, one could use the value that maximizes the GICC in this data set (see Figure 5).

6. Conclusion

In this paper, we propose the concept of the graphical intra-class correlation coefficient using multivariate probit mixed-linear models. The GICC is defined as $\frac{tr(\Sigma_x)}{tr(\Sigma_x)+D}$. We used a Monte Carlo EM algorithm to obtain the MLE of Σ_x, while a Gibbs sampler was used in the E-step. We show the results of GICC in varied simulation settings and in the KIRBY21 test-retest datasets.

While providing GICC, the estimation procedure can also be generalized to multivariate probit mixed-linear model with fixed and random covariates components, which is:

$$o_{ij}(d)=\mathbf{I}_{(y_{ij}(d)>0)}\ y_{ij}(d)=\sum_{p=1}^{P}\mu_{ip}(d)\beta_p(d)+\sum_{r=1}^{R}x_{ir}(d)\eta_{ir}(d)+u_{ij}(d)\ \eta_{ir}\sim\mathbf{N}(\mathbf{0},\mathbf{\Sigma}_r),i.i.d.\ \mathbf{u}_{ij}\sim\mathbf{N}(\mathbf{0},\mathbf{I}),i.i.d..$$

In the EM algorithm, η and $\mathbf{y}$ can be treated as full data and the procedure in Section 3 follows. In section 3.3, we also calculate the observed information matrix for the fixed effects which can provide confidence intervals for β's. Moreover, the procedure can also be used multivariate generalized mixed-linear models, such as multivariate poisson or logistic regression.

Currently, our method works for small graphs. As the number of nodes in a graph increases, the number of parameters of interests grows quadratically ($D \sim O(N^2)$ and $\#\{\sigma_{ij}\} \sim O(D^2)$). Thus a graph with 100 nodes will have tens of millions of parameters to estimate. The Gibbs sampler could not be implemented effectively in such cases. Therefore, the algorithm currently requires a relatively small number of nodes for each graph (typically less than 10). In order to achieve faster convergence rate as well as control the Monte Carlo error induced by Gibbs sampler, ascent based MCEM [27] and acceleration EM algorithm ([28]) could be implemented.

Notice that from the application, GICC could also be used for choosing thresholds for dichotomizing raw graphs. The value that maximizes the GICC is a reasonable threshold, since it yields the best reproducibility of a well known benchmark data set.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In summary, GICC provides us a way to measure the reproducibility of repeated graphical measurements. The current algorithm gives us the estimates of GICC for relatively small graphs. To our knowledge, GICC for large graphs has not been addressed before and therefore deserves further investigation.

References

1. Fisher, RA.; Genetiker, S.; Fisher, RA.; Genetician, S.; Britain, G.; Fisher, RA.; Généticien, S. Statistical methods for research workers. Vol. 14. Oliver and Boyd Edinburgh; 1970.
2. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. Psychological reports. 1966; 19(1):3–11. [PubMed: 5942109]
3. Shrout PE, Fleiss JL, et al. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979; 86(2):420–428. [PubMed: 18839484]
4. Stanish WM, Taylor N. Estimation of the intraclass correlation coefficient for the analysis of covariance model. The American Statistician. 1983; 37(3):221–224.
5. Di C-Z, Crainiceanu CM, Caffo BS, Punjabi NM. Multilevel functional principal component analysis. Annals of Applied Statistics. 2009; 3(1):458–488. [PubMed: 20221415]
6. Shou H, Eloyan A, Lee S, Zipunnikov V, Crainiceanu A, Nebel M, Caffo B, Lindquist M, Crainiceanu C. Quantifying the reliability of image replication studies: The image intraclass correlation coefficient (i2c2). Cognitive, Affective, & Behavioral Neuroscience. 2013; 13(4):714–724.
7. Di Martino A, Scheres A, Margulies D, Kelly A, Uddin L, Shehzad Z, Biswal B, Walters J, Castellanos F, Milham M. Functional connectivity of human striatum: a resting state fmri study. Cerebral cortex. 2008; 18(12):2735–2747. [PubMed: 18400794]
8. Guye M, Bettus G, Bartolomei F, Cozzone PJ. Graph theoretical analysis of structural and functional connectivity mri in normal and pathological brain networks. Magnetic Resonance Materials in Physics, Biology and Medicine. 2010; 23(5–6):409–421.
9. Huang S, Li J, Sun L, Ye J, Fleisher A, Wu T, Chen K, Reiman E. Learning brain connectivity of alzheimer's disease by sparse inverse covariance estimation. NeuroImage. 2010; 50(3):935–949. [PubMed: 20079441]
10. Salvador R, Suckling J, Coleman MR, Pickard JD, Menon D, Bullmore E. Neurophysiological architecture of functional magnetic resonance images of human brain. Cerebral Cortex. 2005; 15(9):1332–1342. [PubMed: 15635061]
11. Van Den Heuvel MP, Hulshoff Pol HE. Exploring the brain network: a review on resting-state fmri functional connectivity. European Neuropsychopharmacology. 2010; 20(8):519–534. [PubMed: 20471808]
12. Ridout MS, Demetrio CG, Firth D. Estimating intraclass correlation for binary data. Biometrics. 1999; 55(1):137–148. [PubMed: 11318148]
13. Rodrıguez G, Elo I. Intra-class correlation in random-effects models for binary data. The Stata Journal. 2003; 3(1):32–46.
14. Zou G, Donner A. Confidence interval estimation of the intraclass correlation coefficient for binary outcome data. Biometrics. 2004; 60(3):807–811. [PubMed: 15339305]
15. Zager LA, Verghese GC. Graph similarity scoring and matching. Applied mathematics letters. 2008; 21(1):86–94.
16. Blondel VD, Gajardo A, Heymans M, Senellart P, Van Dooren P. A measure of similarity between graph vertices: Applications to synonym extraction and web searching. SIAM review. 2004; 46(4): 647–666.
17. Chan JS, Kuk AY. Maximum likelihood estimation for probit-linear mixed models with correlated random effects. Biometrics. 1997:86–97.
18. McCulloch CE. Maximum likelihood variance components estimation for binary data. Journal of the American Statistical Association. 1994; 89(425):330–335.
19. Horrace WC. Some results on the multivariate truncated normal distribution. Journal of Multivariate Analysis. 2005; 94(1):209–221.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

20. Kotecha, JH.; Djuric, PM. Gibbs sampling approach for generation of truncated multivariate gaussian random variables. Acoustics, Speech, and Signal Processing, 1999; Proceedings., 1999 IEEE International Conference on; 1999. p. 1757-1760.IEEE
21. Wilhelm S, MBG. tmvtnorm: Truncated Multivariate Normal and Student t Distribution. r package version 1.4-8. 2013 URL http://CRAN.R-project.org/package=tmvtnorm.
22. Louis TA. Finding the observed information matrix when using the em algorithm. Journal of the Royal Statistical Society. Series B (Methodological). 1982:226–233.
23. Deuker L, Bullmore ET, Smith M, Christensen S, Nathan PJ, Rockstroh B, Bassett DS. Reproducibility of graph metrics of human brain functional networks. Neuroimage. 2009; 47(4): 1460–1468. [PubMed: 19463959]
24. Telesford QK, Morgan AR, Hayasaka S, Simpson SL, Barret W, Kraft RA, Mozolic JL, Laurienti PJ. Reproducibility of graph metrics in fmri networks. Frontiers in neuroinformatics. 4
25. Telesford QK, Burdette JH, Laurienti PJ. An exploration of graph metric reproducibility in complex brain networks. Frontiers in neuroscience. 7
26. Landman BA, Huang AJ, Gifford A, Vikram DS, Lim IAL, Farrell JA, Bogovic JA, Hua J, Chen M, Jarso S, et al. Multi-parametric neuroimaging reproducibility: a 3-t resource study. Neuroimage. 2011; 54(4):2854–2866. [PubMed: 21094686]
27. Caffo BS, Jank W, Jones GL. Ascent-based monte carlo expectation– maximization. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2005; 67(2):235–251.
28. Varadhan R, Roland C. Simple and globally convergent methods for accelerating the convergence of any em algorithm. Scandinavian Journal of Statistics. 2008; 35(2):335–353.

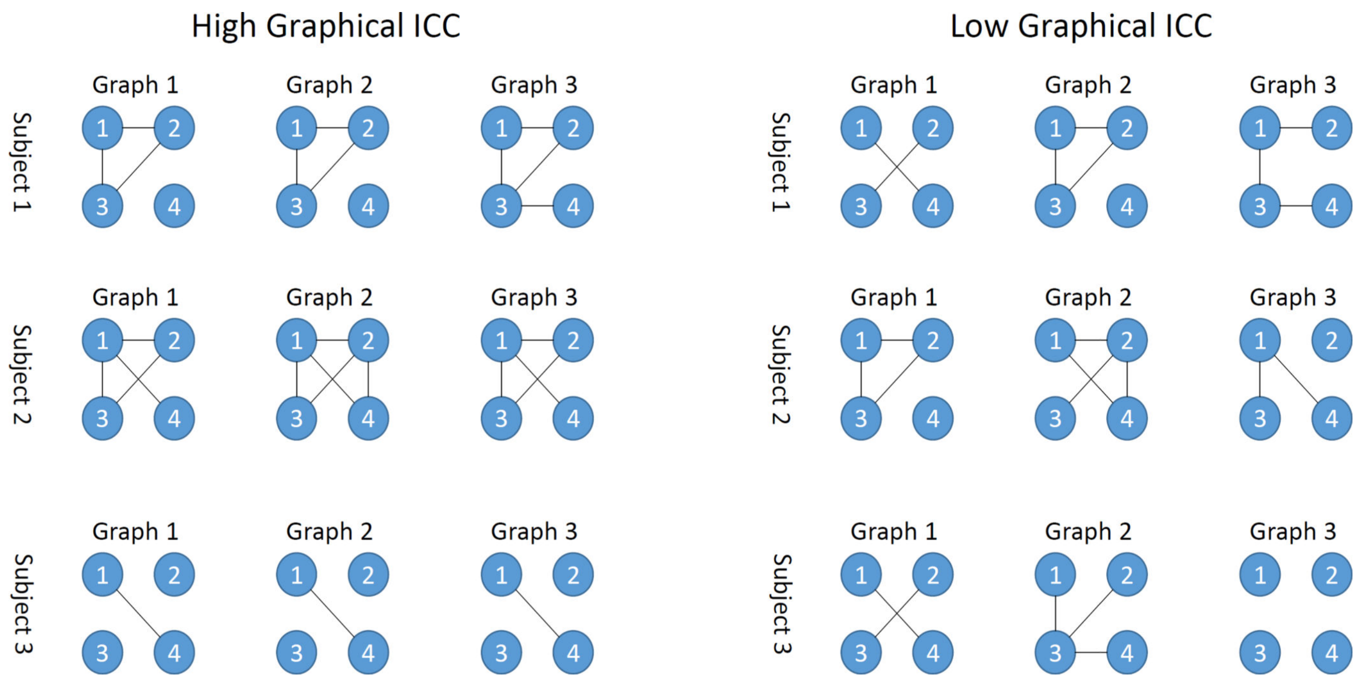

Figure 1.
The left panel shows a high GICC case, where graphical measurements are similar within subjects. The right panel illustrates a low GICC case, where graphical measurements are less consistent within subjects.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

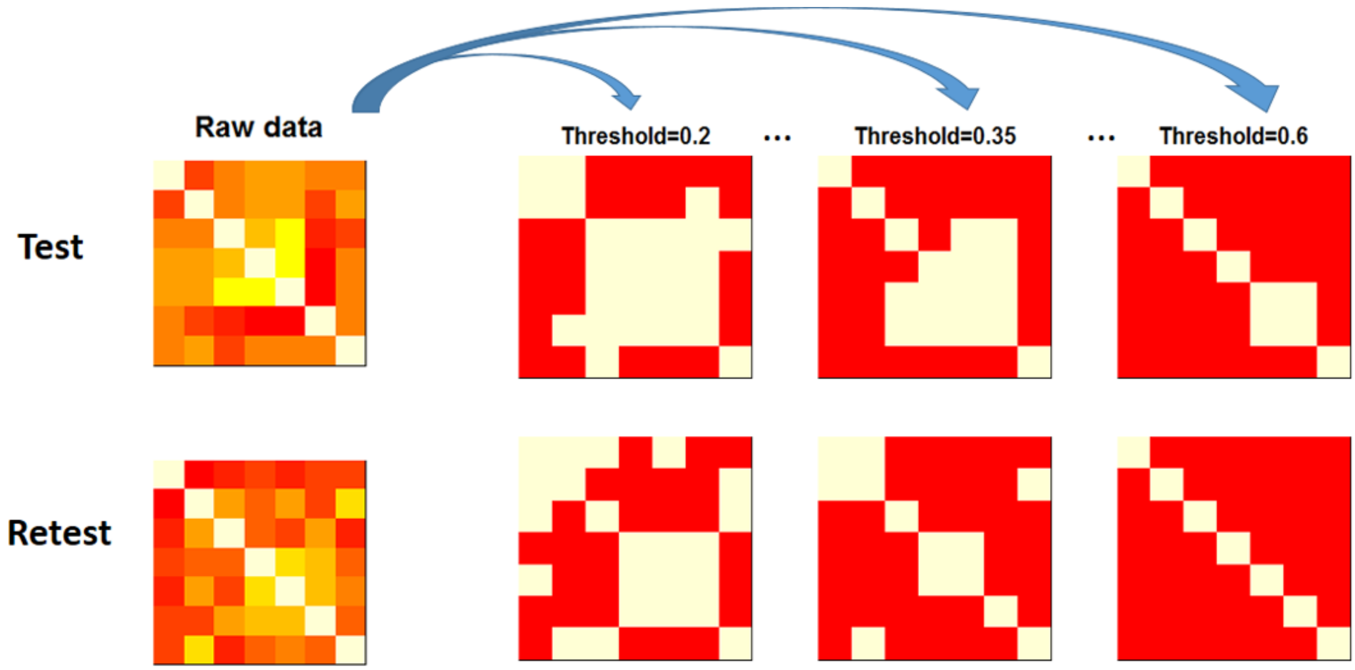

Figure 2.

The figure illustrates two repeated measurements for one subject. On the left, raw correlations between seven nodes are illustrated. Then the raw correlations are dichotomizing using different thresholds (0.2, 0.35, 0.6 are listed here). Our algorithm is then implemented on binary graphes using each threshold. Red suggests lower value and white (yellow) suggests higher value. In the binary graph on the right, red indicates 1 and yellow indicates 0.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

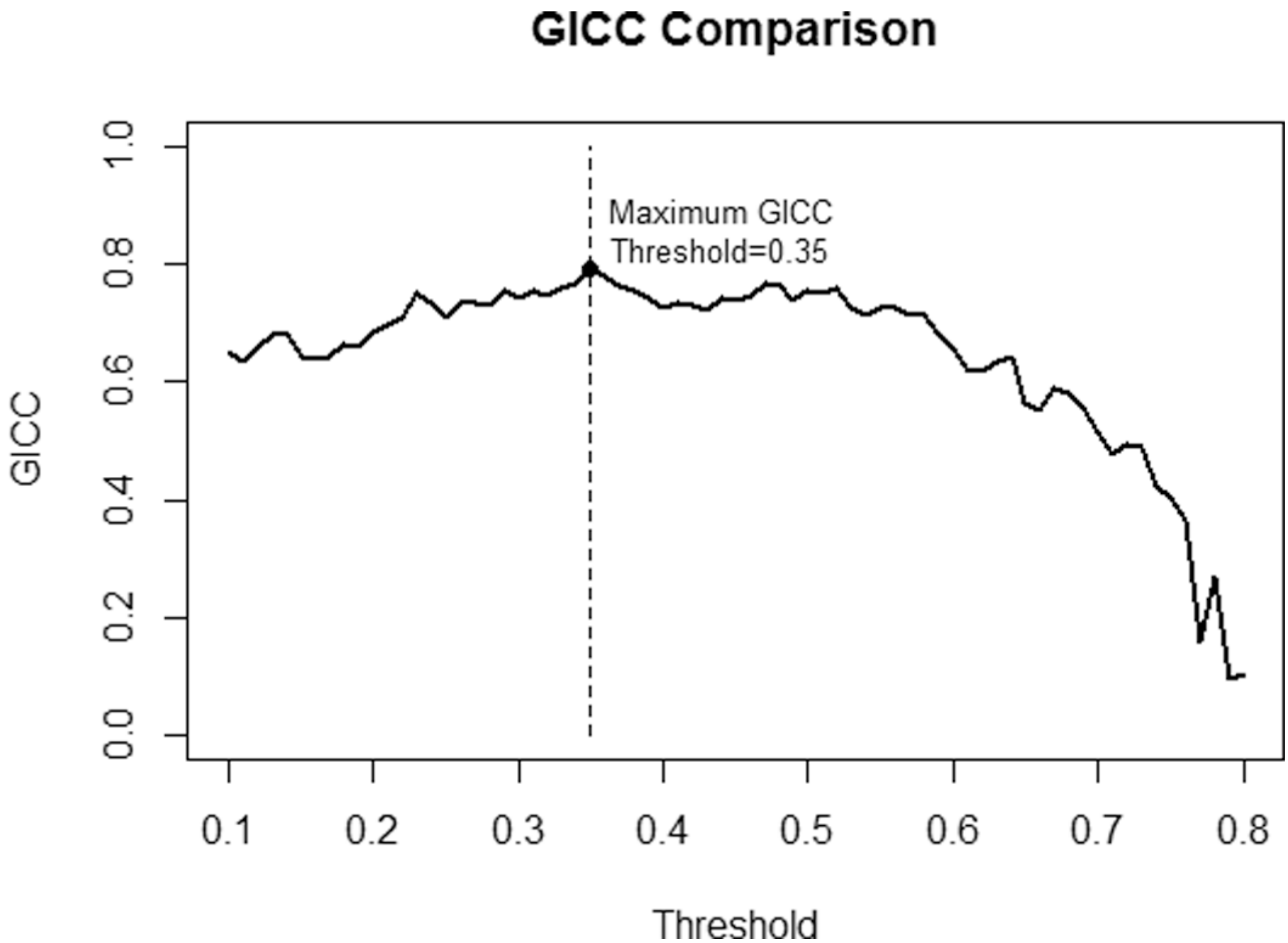

Figure 3.

The calculated GICC under different thresholds. The threshold were picked equally spaced from 0.1 to 0.8 using grid 0.01. The maximized GICC is indicated in the figure, which corresponds to a 0.35 threshold.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Simulation results

Setting	Estimates for σ_{ii}					ICC est.
	$\sigma_{1,1}$	$\sigma_{2,2}$	$\sigma_{3,3}$	$\sigma_{4,4}$	$\sigma_{5,5}$	*ICC*
$I = 100, J = 2, r = 2$	2.37 (1.18)	2.48 (1.09)	2.38 (1.11)	2.35 (1.04)	2.34 (1.01)	0.702
	$\sigma_{6,6}$	$\sigma_{7,7}$	$\sigma_{8,8}$	$\sigma_{9,9}$	$\sigma_{10,10}$	
$ICC_{true} = 2/3$	2.45 (1.09)	2.36 (1.06)	2.43 (1.13)	2.43 (1.04)	2.38 (1.18)	(0.033)
	$\sigma_{1,1}$	$\sigma_{2,2}$	$\sigma_{3,3}$	$\sigma_{4,4}$	$\sigma_{5,5}$	*ICC*
$I = 100, J = 2, r = 4$	4.56 (2.11)	4.74 (2.20)	4.62 (2.30)	4.40 (2.01)	4.56 (2.28)	0.817
	$\sigma_{6,6}$	$\sigma_{7,7}$	$\sigma_{8,8}$	$\sigma_{9,9}$	$\sigma_{10,10}$	
$ICC_{true} = 4/5$	4.65 (2.12)	4.63 (2.04)	4.52 (2.07)	4.61 (2.04)	4.51 (2.31)	(0.025)
	$\sigma_{1,1}$	$\sigma_{2,2}$	$\sigma_{3,3}$	$\sigma_{4,4}$	$\sigma_{5,5}$	*ICC*
$I = 100, J = 4, r = 2$	2.02 (0.60)	2.10 (0.64)	2.07 (0.61)	2.04 (0.65)	2.10 (0.65)	0.672
	$\sigma_{6,6}$	$\sigma_{7,7}$	$\sigma_{8,8}$	$\sigma_{9,9}$	$\sigma_{10,10}$	
$ICC_{true} = 2/3$	2.08 (0.59)	2.06 (0.60)	2.08 (0.63)	2.08 (0.65)	2.05 (0.61)	(0.026)
	$\sigma_{1,1}$	$\sigma_{2,2}$	$\sigma_{3,3}$	$\sigma_{4,4}$	$\sigma_{5,5}$	*ICC*
$I = 100, J = 4, r = 4$	4.00 (1.29)	4.08 (1.36)	4.04 (1.28)	3.96 (1.29)	4.17 (1.36)	0.800
	$\sigma_{6,6}$	$\sigma_{7,7}$	$\sigma_{8,8}$	$\sigma_{9,9}$	$\sigma_{10,10}$	
$ICC_{true} = 4/5$	4.04 (1.24)	4.04 (1.20)	4.10 (1.34)	4.09 (1.32)	4.05 (1.24)	(0.020)
	$\sigma_{1,1}$	$\sigma_{2,2}$	$\sigma_{3,3}$	$\sigma_{4,4}$	$\sigma_{5,5}$	*ICC*
$I = 200, J = 2, r = 2$	2.08 (0.68)	2.22 (0.77)	2.23 (0.74)	2.17 (0.73)	2.19 (0.70)	0.683

Setting	Estimates for σ_{ii}					ICC est.
	$\sigma_{6,6}$	$\sigma_{7,7}$	$\sigma_{8,8}$	$\sigma_{9,9}$	$\sigma_{10,10}$	
$ICC_{true} = 2/3$	2.17 (0.76)	2.17 (0.68)	2.16 (0.71)	2.20 (0.77)	2.13 (0.74)	(0.026)
	$\sigma_{1,1}$	$\sigma_{2,2}$	$\sigma_{3,3}$	$\sigma_{4,4}$	$\sigma_{5,5}$	*ICC*
$I = 200, J = 2, r = 4$	4.07 (1.38)	4.33 (1.41)	4.25 (1.57)	4.20 (1.49)	4.18 (1.42)	0.806
	$\sigma_{6,6}$	$\sigma_{7,7}$	$\sigma_{8,8}$	$\sigma_{9,9}$	$\sigma_{10,10}$	
$ICC_{true} = 4/5$	4.20 (1.52)	4.20 (1.36)	4.25 (1.43)	4.30 (1.47)	4.15 (1.47)	(0.020)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Compare to Benchmark

	Models			
	ave. ICC(1)	**ICC(1) for vec.mean**	**1-var probit**	**GICC**
$ICC_{true} = 0.667$	0.457 (0.028)	0.812 (0.035)	0.613 (0.025)	0.696 (0.029)

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3

Running Time (in seconds)

	I=50			I=100		
	D = 10	**D = 20**	**D = 30**	**D = 10**	**D = 20**	**D = 30**
J = 2	17.2	28.7	62.4	33.5	57.2	124.8
J = 3	20.7	61.5	215.4	41.3	122.8	430.6
J = 4	28.0	144.5	586.5	55.9	289.0	1173.3