Towards a psychology of individuals: the ergodicity information index and a bottom-up approach for finding generalizations

Hudson Golino[1], John Nesselroade[1], & Alexander P. Christensen[2]

[1] Department of Psychology, University of Virginia

[2] Department of Psychology and Human Development, Peabody College, Vanderbilt University

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

The authors made the following contributions. Hudson Golino: Conceptualization, Methodology, Formal Analysis, Writing - Original Draft Preparation, Writing - Review & Editing; John Nesselroade: Writing - Original Draft Preparation, Writing - Review & Editing; Alexander P. Christensen: Writing - Original Draft Preparation, Writing - Review & Editing.

Correspondence concerning this article should be addressed to Hudson Golino, 485 McCormick Road, Gilmer Hall, Room 102, Charlottesville, VA 22903. E-mail: hfg9s@virginia.edu

Abstract

In the last half of the 20th century, psychology and neuroscience have experienced a renewed interest in intraindividual variation. To date, there are few quantitative methods to evaluate whether a population structure (between-person) is likely to hold for individual people. We present a network information theoretic approach to evaluate the extent to which a system possesses the ergodic property. We introduce a new metric, the ergodicity information index (EII), that can inform whether a set of multivariate time series (or a set of intensive longitudinal measures) should be represented as multiple individual structures or as a single population structure. The EII index quantifies the amount of information lost by representing all individuals with a single population structure. If the individuals don't have a similar structure, representing them with a single population network leads to a loss of information. The EII index value will be higher than in cases where all individuals have a similar structure. A Monte-Carlo simulation is implemented to test the EII index, and the results show that the new index has a 94% accuracy in differentiating data in which all individuals have a similar structure vs. data in which the individuals don't have a similar structure. The paper also presents two new techniques designed to help applied researchers to analyze data when the ergodicity property does not hold. The EII bootstrap test obtains a sampling distribution of EII values as if all participants in the data have the population network structure and compare this null distribution to the empirical EII value. Significant differences indicate that the empirical data cannot be expected to be generated from an ergodic process, and the population structure is not sufficient to describe all individuals. The Domenico clustering method estimates the Von Neumann entropy of two networks and computes their Jensen-Shannon Distance (JSD). Then, a complete-linkage agglomerative hierarchical clustering method is applied to the JSD, and the clustering partitioning is obtained via modularization maximization. The Domenico clustering method allows the discovery of groups of individuals with a similar structure. Finally, two empirical examples are shown, one using data from an intensive longitudinal experience sampling study

examining Big Five personality measured by the Big Five Inventory-2, and the other using resting state neuroimaging data taken from a study examining creativity which used the 268-node Shen brain atlas. Starting with personality, the bootstrap EII test was significant, suggesting that the BFI-2 data were nonergodic. Following up on the bootstrap EII test, the information clustering was applied, and the single cluster test was performed. The single cluster test suggested that the empirical networks had significantly greater JSD values than the random networks meaning that the single cluster detected was not meaningful, and each individual in the sample is unique. The brain networks had a significant bootstrap EII test and significantly larger JSD values than random networks. These results, in line with the personality data, suggest that (resting state) brain networks are not ergodic, and no meaningful groups can be formed (i.e., each individual is unique). In sum, the personality and (resting state) brain networks do not possess ergodicity and therefore lose information when data are aggregated into a single population.

Keywords: keywords

Word count: X

Towards a psychology of individuals: the ergodicity information index and a bottom-up approach for finding generalizations

---

Introduction

In the last half of the 20th century, psychology has experienced a renewed interest in within-person (intraindividual) variation. One principal line of development sprang from the introduction of P-technique factor analysis (Cattell, Cattell, & Rhymer, 1947). P-technique involves the periodic collection (e.g. hourly, daily) of responses to a battery of measures taken from a single person as opposed to across multiple people. The variation and covariation of these measures are within a person (intraindividual) as opposed to the more common between-person variation and covariation (interindividual) studied by differential psychology. Although P-technique did not quickly become a standard tool for exploring variability, its promise and the promise of studying intraindividual variation has attracted increasing attention (Bereiter, 1963; Molenaar, 2004; Nesselroade & Ford, 1985). Today, there is a considerable amount of research collecting data on intraindividual variability (Beck & Jackson, 2022; e.g., ; Diehl, Hooker, & Sliwinski, 2014; Fisher, Reeves, Lawyer, Medaglia, & Rubel, 2017; Gomes & Golino, 2015; Hultsch, Strauss, Hunter, & MacDonald, 2008; Ram, Gerstorf, Lindenberger, & Smith, 2011), clearly having an impact on the field.

The re-emergence of within-person research, in an arena long dominated by between-person research, is not surprising. The accessibility of new technologies that are able to capture intraindividual variability (e.g., smartphones) has enabled researchers unparalleled opportunity to study a person rather than people. Intraindividual and interindividual variability has often been cast in terms of the idiographic versus nomothetic debate (see e.g., Lamiell, 1998). The recognition of these important thrusts has several consequences. There is an obvious need to formalize the definitions of both conceptions, and

to do it in such a way that they can be readily distinguished at the conceptual level while also highlighting their relation to each other (e.g., Molenaar, Huizenga, & Nesselroade, 2003; Oertzen, Schmiedek, & Voelkle, 2020; Schmiedek, Lövdén, Oertzen, & Lindenberger, 2020; Wright & Zimmermann, 2019).

Accompanying that effort, ways to distinguish between, and capitalize on the two kinds of variation in empirical data must also be created and made operational. For example, to what extent can we expect to characterize the way individuals differ from each other with the same structures that characterizes how a person changes over time? This question harks back to the old debate in developmental psychology about the relative merits of cross-sectional versus longitudinal research designs and whether cross-sectional data can be relied on to furnish accurate information regarding change over time at the individual level.

More recently, the ergodic property has surfaced in this context as a condition under which one can expect within-person and between-person structures to match (Fisher, Medaglia, & Jeronimus, 2018). If a system's process is ergodic, between-person structures can be used to represent within-person structures (Molenaar et al., 2003) but, among developmental psychologists at least, there seems to be little reason to think that ergodicity will be a property of many (if any) developmental processes. To date, there are few quantitative methods to evaluate whether a between-person structure is likely to hold for individual people—that is, whether the system possesses the ergodic property. In the present research, we present an information theoretic approach to evaluate the extent to which a system possesses the ergodic property. From an information theoretic perspective, the ergodic property can be framed in terms of the amount of information lost representing a set of measures as a single between-person structure (nomothetic structure) instead of as multiple within-person structures (within-person and idiographic structures).

Techniques to analyze complex systems with dynamic interactions between variables have a long history in statistical physics (e.g., Jaynes, 1957) and psychology (Boker, 2018;

Cattell, 1965; Guttman, 1953), culminating in modern approaches such as network science (Epskamp & Fried, 2018; Epskamp, Waldorp, Mõttus, & Borsboom, 2018a). In networks, variables are represented as nodes (circles) and the edges (lines) between the nodes represent associations between variables. There are many ways to characterize complex networks (Newman, 2010), from the type of edge (e.g., directed or undirected, weighted or unweighted) to the complexity of the network (e.g., algorithmic complexity; Morzy, Kajdanowicz, & Kazienko, 2017; Zenil, Kiani, & Tegnér, 2018). Networks are a common method for representing complex systems and the interactions between their components.

We introduce a new metric – termed *ergodicity information index* (EII) – that can inform whether a set of variables should be represented as multiple individual structures (multiple individual networks, multiplex networks) or as a single, between-individual structure (unique single network, aggregate or population network). The EII characterizes the relative algorithm complexity of the population structure with regard to multiple individual networks taking into consideration the number of underlying dimensions (e.g., communities, latent factors). Algorithm complexity of multiplex networks can be used to determine the optimal number of layers needed to represent a multiplex network and to detect structural and dynamical similarities among their layers (Santoro & Nicosia, 2020). The representation of intraindividual structures as multiplex networks and the quantification of their information relative to a single, population network structure are two central ideas in the development of our EII.

The paper is organized as follows. The first section briefly introduces how intraindividual and interindividual structures can be estimated in a single framework using dynamic exploratory graph analysis (DynEGA; Golino, Christensen, Moulder, Kim, & Boker, 2022), an approach that combines generalized local linear approximation (Boker, Deboek, Edler, & Keel, 2010) and exploratory graph analysis (Golino & Epskamp, 2017; Golino, Shi, et al., 2020) to estimate dynamical factors in (intensive) longitudinal measures at different levels of analysis (individual, group, and/or population). In this section, we reframe the

within- and between-person problem from a network perspective in which the intraindividual structures are represented as a multiplex network (i.e., a collection of individual networks), and the interindividual structure as a single (population) network.

In the second section, we develop the EII, introducing information theoretic concepts such as algorithm complexity that are necessary to define its meaning and interpretation. Next, a Monte Carlo simulation study is implemented to investigate the accuracy of the EII to differentiate between within- and between-person structures. Afterwards, we introduce an information theoretic approach to clustering, which can be used if the system is determined to be nonergodic.

Finally, one synthetic and two empirical examples from personality and neuroscience will be used to demonstrate how the EII can be used to determine whether a system possesses the ergodic property, and determine ergodic clusters (or groups) when the system is not. These new techniques are a step forward in the psychology of individuals, enabling the identification of generalizable constructs using a bottom-up approach (from individuals to group of individuals with common network characteristics).

Representing intraindividual and interindividual structures as networks

To reframe the within- vs. between-person problem using a information theoretic network approach, we first need to show how networks can be estimated in (intensive) longitudinal data in a way that can generate both multiplex networks for the individuals (i.e., multiple individual, within-person networks) and a single population or between-person network. Recently, Golino et al. (2022) introduced a technique termed *dynamic exploratory graph analysis* (**DynEGA**), combining techniques from dynamical systems (i.e., time-delay embedding and generalized local linear approximation; Boker et al., 2010) and exploratory graph analysis (Golino & Epskamp, 2017; Golino, Shi, et al., 2020) – a network psychometric approach for dimensionality assessment and reduction. The *DynEGA* technique can be used

to estimate dynamical communities (e.g., latent factors) in (intensive) longitudinal measures at different levels of analysis (individual, group and/or population).

Network models have been proposed in psychological research for decades (e.g.,; Boker, 2018; Cattell, 1965; Guttman, 1953). More recently, a number of developments in network modeling in psychology originated a new area termed *network psychometrics* (Epskamp, 2018) that relies mostly on the Gaussian graphical model (GGM; Lauritzen, 1996). Assuming multivariate normality, a GGM can be obtained by modeling the inverse of the variance-covariance matrix (and standardizing it to obtain partial correlations) in a way that non-zero elements are freely estimated (Epskamp et al., 2018a), generating a sparse model of the variance-covariance matrix (Epskamp, Rhemtulla, & Borsboom, 2017). As Epskamp et al. (2018a) note, inverting and standardizing the variance-covariance matrix won't lead to partial correlations that are exactly zero, meaning that the GGM is saturated. Regularization techniques, such as a variant of the *least absolute shrinkage and selection operator* (LASSO; Tibshirani, 1996) termed *graphical LASSO* (GLASSO; Friedman, Hastie, & Tibshirani, 2008), are generally used in network psychometrics (Epskamp & Fried, 2018; see; Epskamp et al., 2018a).

The GLASSO is a technique that is very fast to estimate both the model structure and the parameters of a sparse GGM (Epskamp, Waldorp, Mõttus, & Borsboom, 2018b). It has a tuning parameter (γ), that can be chosen in a way to minimize the extended Bayesian information criterion (EBIC; Chen & Chen, 2008), which is used to estimate optimal model fit and has been shown to accurately retrieve the true network structure in simulation studies (Epskamp & Fried, 2018; Foygel & Drton, 2010; Williams & Rast, 2020; Williams, Rhemtulla, Wysocki, & Rast, 2019).

The most used network estimation approach in psychological research (termed *EBICglasso*) is implemented in the `qgraph` package (version 1.4.1; Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012). The `EBICglasso` function (Epskamp et al.,

2012) samples 100 logarithmically-spaced values of λ, following Foygel and Drton (2010). The ratio range of λ can be set by the user (defaults to 0.01). γ controls the severity of the model selection (defaults to 0.50). EBIC is computed for values of `gamma` larger than zero, and the value of λ that minimizes this information criteria is selected, generating a network with regularized edges.

Network psychometrics on cross-sectional data has progressed into dimensionality assessment, where networks are estimated using the GGM or other network techniques (see Golino, Shi, et al., 2020) and an algorithm for detecting communities in weighted networks (Walktrap; Pons & Latapy, 2005) is used to identify latent factors. Golino and Epskamp (2017) called this approach *exploratory graph analysis* (**EGA**). Simulation studies have found EGA to perform as well as the most accurate factor analytic method, parallel analysis, and produce the best large-sample properties of all the methods evaluated (Golino & Epskamp, 2017; Golino, Shi, et al., 2020).

In (intensive) longitudinal data, EGA can be used, but instead of using the variance-covariance matrix of the raw data, it uses the variance-covariance of m-order derivatives. The resulting network structure (GGM of the m-order derivatives) conveys information on how variables are *changing* over time. As a result, the communities identified using the Walktrap algorithm reflect not simple, static factors, but dynamical factors of nodes that are fluctuating similarly as a function of time. Golino et al. (2022) called this technique the *dynamic exploratory graph analysis* (*DynEGA*), merging network psychometrics, dynamical systems modeling, and dimensionality assessment into a single framework that can estimate structures at the individual or group/population levels.

DynEGA starts by transforming the time series of each variable $V = \{v_1, v_2, ..., v_N\}$ into a time delay embedding matrix $\mathbf{X}^{(n)}$, where n is the number of embedding dimensions. As pointed by Golino et al. (2022), a time delay embedding matrix is used to reconstruct the attractor of a dynamical system using a single sequence of observations (Takens, 1981;

Whitney, 1936), preserving the phase-space dynamics of the system. In the time delay embedding matrix, each row is a phase-space vector (Rosenstein, Collins, & De Luca, 1993):

$$X = [X_1\ X_2\ ...\ X_M]', \tag{1}$$

where X_i is the state of the system at discrete time i and is given by:

$$X_i = \left[x_i\ x_{i+\tau}\ ...\ x_{i+(n-1)\tau}\right], \tag{2}$$

where τ is the number of observations to offset successive embeddings (i.e., lag or reconstruction delay) and n is the embedding dimension. The time-delay embedding matrix is a $M \times n$ matrix, where $M = N - (n-1)\tau$ and N is the number of time points.

Once the time series of each variable collected in an (intensive) longitudinal study and transformed into a time-delay embedding matrix $\mathbf{X}^{(n)}$, derivatives can be estimated using generalized local linear approximation (Boker et al., 2010; **GLLA**; Deboeck, Montpetit, Bergeman, & Boker, 2009). Derivatives can represent different aspects of change such as the rate of change or velocity at which the variable is changing over time (first-order derivatives) and the speed of the rate of change or acceleration (second-order derivatives).

Deboeck et al. (2009) and Boker et al. (2010) show how derivatives can be estimated in the GLLA framework:

$$\mathbf{Y} = \mathbf{XL}(\mathbf{L}'\mathbf{L})^{-\mathbf{1}}, \tag{3}$$

where $\mathbf{Y}$ is a matrix of derivative estimates, $\mathbf{X}$ is a time delay embedding matrix (with n embedding dimensions; to simplify the notation, $\mathbf{X} = \mathbf{X}^{(n)}$), and $\mathbf{L}$ is a matrix with the weights expressing the relationship between the embedding matrix and the derivative estimates. The weight matrix $\mathbf{L}$ is a $n \times \alpha$ matrix, where n is the number of embedding

dimensions and α is the (maximum) order of the derivative. Each column of the weight matrix is estimated as follows, considering the order of the derivatives going from zero to k, $\alpha = [0, 1, ..., k]$:

$$\mathbf{L}_\alpha = \frac{[\Delta_t(v - \bar{v})]^\alpha}{\alpha!} \tag{4}$$

where Δ_t is the time between successive observations in the time series, v is a vector from one to the number of embedded dimensions (i.e., $v = [1, 2, ...n]$), $\bar{v}$ is the mean of v, α is the order of the derivative of interest, and $\alpha!$ is the factorial of α.

After estimating the derivatives for all time series (i.e., all variables), EGA is used to estimate the multiplex networks (intraindividual or within-person structures). In this process, each matrix of derivatives will generate a different network and dimensionality structure (represented by clusters of nodes in the network) for each individual. The population (between-person or nomothetic) structure can be estimated by stacking the derivative matrix of each individual (i.e., row-binding the matrices) and applying EGA to the stacked matrix. In both cases (i.e., individual or population structures), the resulting clusters in the networks corresponds to variables that are changing together (Golino et al., 2022).

An example illustrates how this technique works: Suppose we ask two people to answer eight items of depression once a day for 100 days. After applying *DynEGA* to the data, we would obtain three important types of information (see Figure 1): the derivatives for the eight variables, the network structure for each person (intraindividual), and the network structure of the two people combined (interindividual). The left side of Figure 1 shows the time series of first-order derivatives while the network structure of each individual are at the center (top and bottom represents subject one and two, respectively). The interindividual network structure is depicted in the right side of Figure 1. As can be seen in the network structure of person one (top network at the center of Figure 1), the pair of nodes pink and green are not connected (i.e. there is no edge linking these pairs of nodes). However, the

same pair of nodes are connected in the network structure of person two (bottom network at the center of Figure 1), and in the between-individual network structure (or population network) depicted at the right side of Figure 1. A question that must be answered now is: how much information is being lost by representing the structure of the two individuals as a single, interindividual network, relative to representing them as a multiplex network? Can the derivatives of person one and two be stacked to estimate a single network or by doing so, is important information about each individual lost? The next section aims to answer these questions and to introduce the *ergodicity information index*.

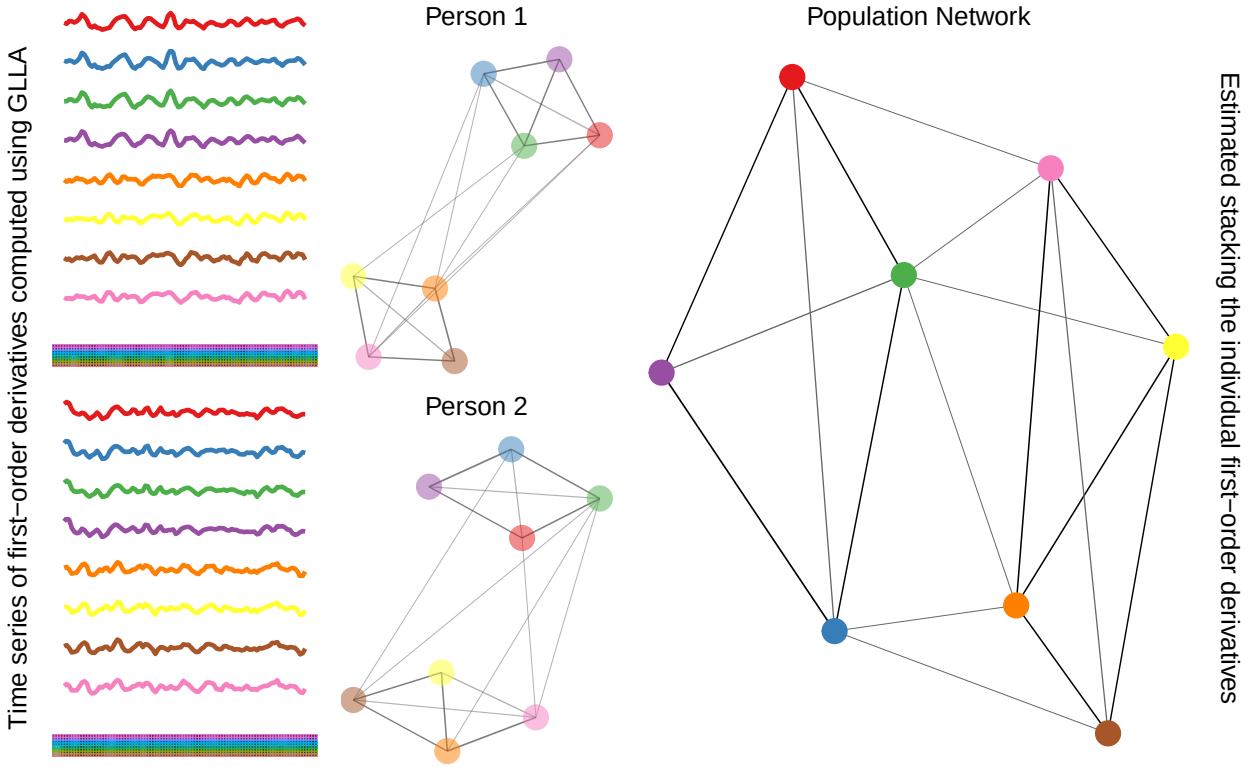

Figure 1

The Ergodicity Information Index

The individual networks and population network in Figure 1 can be compared in terms of their algorithm complexity. Algorithm complexity can be used to analyze complex objects in an unbiased manner using mathematical principles (Zenil et al., 2018), and is based on the

work of Kolmogorov (1968), Martin-Löf (1966), Solomonoff (1964) and others. As Zenil et al. (2018) and Morzy et al. (2017) show, the algorithm (or Kolmogorov) complexity of a string s is formally defined as:

$$K_s = min(|P|, T(P) = s)$$

where P is a program producing the string s when running on a universal Turing machine T, and $|P|$ is the number of bits required to represent P (i.e., the length of P). A Turing machine is a formal model of a general-purpose computer that can be programmed to reproduce any computable object, such as a string (Zenil et al., 2018). The Kolmogorov complexity of a string is defined, in other words, as the length of the shortest possible program that can produce that string as its output (Santoro & Nicosia, 2020). Kolmogorov complexity has a drawback of being incomputable (Zenil et al., 2018), since there is no way to estimate the number of possible programs that could produce the string s (Kolmogorov, 1968). But it can be approximated by using compression algorithms (Morzy et al., 2017; Santoro & Nicosia, 2020), in which the compressed string s is an estimate of K_s.

To obtain an estimation of Kolmogorov complexity in networks, the most common approach is to compute the size of the compressed weighted edge list (Santoro & Nicosia, 2020). For single, unique networks, this is straightforward, but for multiplex networks, the Kolmogorov complexity requires a strategy to encode all individual graphs into a single network. Santoro and Nicosia (2020) proposed the use of a prime-weight encoding matrix Ω that assigns a distinct prime number ($p^{[\alpha]}$) to each individual network (i.e., each of the A layers of the multiplex networks) and sets each element Ω_{ij} equal to the product of the primes associated to the layers where an edge between node i and j exists:

$$\Omega_{ij} = \begin{cases} \prod_{\alpha:\alpha_{ij}^{[\alpha]}=1} p^{[\alpha]} \\ 0 \text{ if } \alpha_{ij}^{[\alpha]} = 0 \; \forall \; \alpha = 1, \ldots, A \end{cases} \tag{5}$$

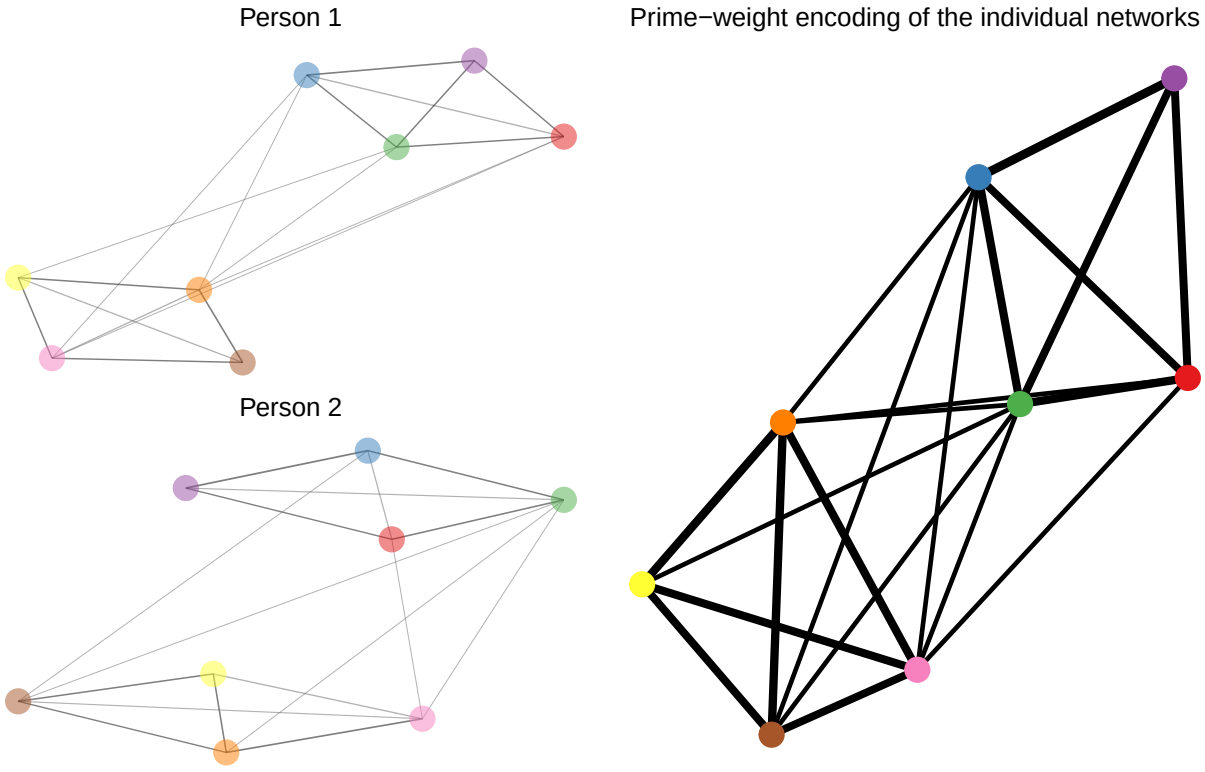

Figure 2

Santoro and Nicosia (2020) proposed a new metric for quantifying the algorithm complexity of multiplex networks that can be computed as the ratio of the (approximate) Kolmogorov complexity of the prime-weight matrix Ω of a multiplex network with A layers and the Kolmogorov complexity of an aggregated network combining all layers.

A similar strategy can be followed to quantify the algorithm complexity of the networks estimated using DynEGA. The multiplex networks are all individual networks estimated using the derivatives computed via generalized local linear approximation in the DynEGA technique. Instead of comparing the algorithm complexity of Ω with an weight aggregation of the multiplex networks, it is more informative to compare it with the

population network (i.e., the network estimated stacking the derivatives estimated using generalized local linear approximation for all individuals). Additionally, in psychology, it is also important to consider the number of latent factors underlying the intensive-longitudinal data. Therefore, our *ergodicity information index* can be computed as:

$$\xi = \sqrt{F_P}\left[\frac{\left(\frac{(K_\Omega)}{K_{P*}}\right)}{\log_{(L_\chi)}}\right]$$

where $\sqrt{F_P}$ is the square-root of the number of factors estimated in the population structure using dynamic exploratory graph analysis, K_Ω is the algorithm complexity of the prime-weight encoding matrix of the individual networks (that composes the multiplex network χ), K_{P*} is the algorithm complexity of the prime-weight transformation of the population network (i.e., each element in the population network, P_{ij}, is transformed such that $P_{ij}* = 2^{P_{ij}}$),and $L\chi$ is the number of distinct edges across the networks that make up the multiplex network (i.e., non-zero edges).

The EII (ξ) computes the amount of information lost representing a set of measures as a single interindividual structure (nomothetic structure) instead of representing the measures as multiple individual structures (within-person or intraindividual structures). Larger values of the ergodicity information index indicate that the intraindividual networks encode a relatively larger amount of information with respect to the population network.

The ratio $\frac{(K_{P*})}{K_\Omega}$ is normalized by the number of distinct edges in the multiplex network (set of individual networks) because a set of networks with larger number of edges is expected to have a higher complexity. Similar to Santoro and Nicosia (2020), we implemented a canonical prime association to compute ξ. In canonical prime association, prime numbers are associated to layers in increasing order of their total number of edges.

In terms of estimation, the Kolmogorov complexity is estimated for each network (i.e., prime-weight transformation of the population network $P*$ and the prime-weight encoding of

the multiplex networks Ω) as the length of the compression of the string formed by their edge list, using the *Gzip* compression algorithm available in the `memCompress` function of base R (R Core Team, 2017). In theory, Kolmogorov complexity can also be computed using other representations of the network such as a string composed by the rows of the Laplacian matrix, degree list, degree distribution, or weights list (Morzy et al., 2017). In this paper, we focus mainly on the algorithm complexity estimated using the weights list, and compare it to the estimation using the edge list. Since Kolmogorov complexity is affected by the order of elements in the edge list, the final estimation of algorithm complexity is based on the mean of 1,000 computations of K over an edge list randomly ordered. One advantage of using Kolmogorov complexity to estimate the complexity of networks is that it is less dependent on the network representation than other metrics of complexity such as entropy-based metrics (Morzy et al., 2017).

The use of the EII implies a different type of ergodicity that we call *super-weak ergodicity*. In a strict definition of ergodicity, there are two central requirements: stationarity for all participants and homogeneity for all time points (Voelkle, Brose, Schmiedek, & Lindenberger, 2014). A softer type of ergodicity termed *weak ergodicity*, requires only that the marginal distributions for all participants and for all time points be identical (Oertzen et al., 2020). The *super-weak ergodicity*, on the other side, doesn't require stationarity for all participants (i.e., the same covariance matrix for all subjects), homogeneity for all time points (i.e., the same covariance matrix across time), or an equal marginal distributions for all participants and time points. It requires a much weaker condition: the algorithm complexity of the population (or between-person) network be similar (but not equal) to the algorithm complexity of the prime-weight encoded network of all individuals.

Suppose we have four people that are assessed using an eight-item questionnaire for 100 days. Persons one and two are more similar than persons three and four, although none of them are exactly equal to one another. DynEGA is used to compute the first-order derivatives for each variable and to generate a network for each individual, and two

between-person or population networks: one for persons one and two, and one for persons three and four. If we calculate the correlation of the derivatives for each variable, none of them are exactly equal to the others. Figure 3 shows the heatmap of the correlation matrices (calculated using the first-order derivatives of each variable) and the resulting network structure for each person and each population with node colors representing the estimated latent factors.

Two factors were identified in individuals one, two and three, and four factors were identified in individual four. The population network for persons one and two indeed shows two factors, while the population network for persons three and four shows four factors, both estimated using DynEGA. Calculating the ergodicity information index (ξ), we obtain $\xi = 1.14$ for individuals one and two and $\xi = 1.29$ for individuals three and four. Therefore, more information is lost by representing the eight measures as a single structure for individuals three and four (bottom of Figure 3, population 2) than for individuals one and two (top of Figure 3, population 1). Another way to interpret the results above is that the intraindividual networks for individuals three and four encode a relatively larger amount of information with respect to the population network, compared to individuals one and two and their population structure.

Simulation Design

To verify the suitability of the ergodicity information index to identify if intensive longitudinal measures should be represented as a set of within-person structures (multiple individual networks) or as a single, between-person or population structure (only one network), a Monte Carlo simulation is implemented. Four data conditions were systematically manipulated: sample size (10, 50 and 100), number of time points (50 and 100), number of variables per factor (4 and 6) and number of factors (2, 3). Two separate set of conditions were used in the simulation. In the first, all individuals had the same number of factors (*Eq* condition). Therefore, representing these individuals using a single population

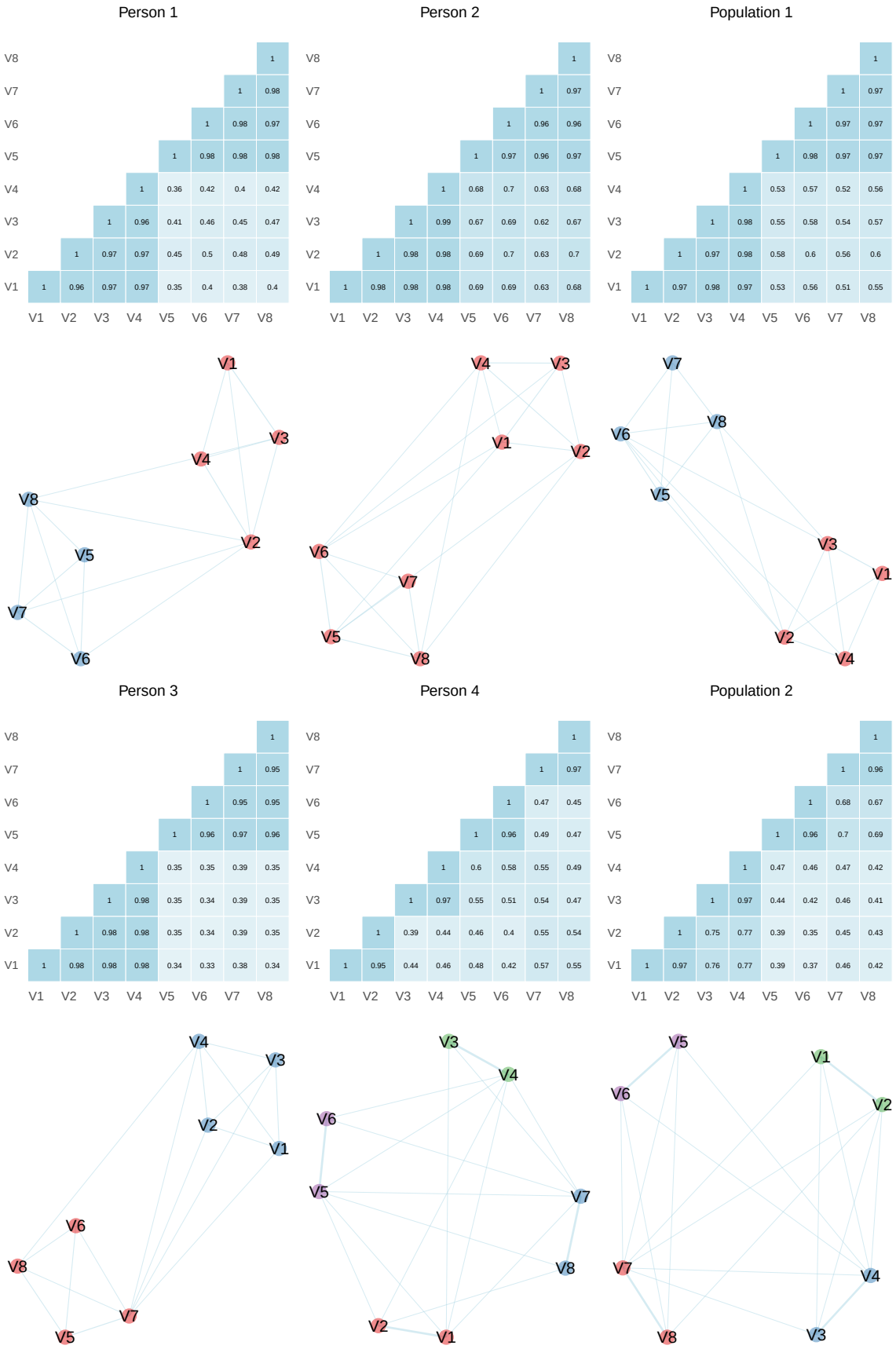
Person 1
Person 2
Population 1
Person 3
Person 4
Population 2
V1
V2
V3
V4
V5
V6
V7
V8

Figure 3

structure is reasonable. In the second set of conditions, half of the subjects had the same number of factors (and variables per factor) as condition one, but the other half had a different configuration, with more or less factors (*NotEq* condition).

The sample sizes were selected to reflect low, moderate, and high samples, consistent with many empirical papers using an intensive longitudinal measurement design that typically don't use data from more than 100 individuals (Liu, Zhou, Palumbo, & Wang, 2016; Schmiedek, Lövdén, & Lindenberger, 2020). Regarding the number of time points and factor loadings (set to unity), they were selected based on the conditions tested by Zhang, Hamaker, and Nesselroade (2008). In terms of the number of variables per factor, three are the minimum required for factor identification (Anderson, 1958). In the present simulation, the number of items were selected to reflect adequate (4) and slightly overindentified (6) factors (Velicer, 1976; Widaman, 1993). The measurement error covariance matrix had 0.1 in the diagonal, so we could see how the EII works under minimum error conditions. This strategy allows for the impact of the other variables systematically manipulated in the simulation to be better understood.

The autoregression coefficients of the DFAS were set to 0.1, 0.4, and 0.8, representing a low, moderate, and high autoregressions. Additionally, two conditions were held constant: the matrix with autoregressive (0.8) and cross-regressive coefficients (0), and the covariance matrix for the random shock (off-diagonal = 0.18; diagonal = 0.36). The values of the cross-regressive coefficients and the random shock matrix were selected following Zhang et al. (2008).

Data Generation

In the Monte Carlo simulation, 500 data matrices were generated for each combination of variables (number of factors, number of variables per factor, number of time points, sample size, autoregression coefficients) according to the DAFS model. First, the matrix of

random shock vectors $\mathbf{v}_t$ was generated following a multivariate normal distribution with mean zeros and $q \times q$ covariance matrix $\mathbf{D}$ (off-diagonal values = 0.18; diagonal values = 0.36), where q is the number of factors and t is the number of time points plus 1,000 (used as the burn-in estimates for the Markov chain). Second, the factor scores are calculated and the first 1,000 estimates are removed (burn-in phase). Third, the measurement error matrix is estimated following a multivariate normal distribution with mean zeros and $p \times p$ covariance matrix Q, where p is the total number of variables (number of variables per factor times F). Finally, the observed variables $\mathbf{Obs}_t$ at time t ($t = 1, 2, ..., N$) are calculated using the following equation 6.

$$\mathbf{Obs}_t = \mathbf{\Lambda}\mathbf{F}_t + \mathbf{e}_t, \tag{6}$$

where $\mathbf{\Lambda}$ is the factor loading matrix ($p \times q$), $\mathbf{F}_t$ is a $q \times 1$ vector of factors at time t, and $\mathbf{e}_t$ is a $p \times 1$ vector with measurement errors following a multivariate normal distribution with mean zeros and covariance matrix Q (Nesselroade, McArdle, Aggen, & Meyers, 2002; Zhang et al., 2008).

The factor scores, $\mathbf{F}_t$, are calculated as follows:

$$\mathbf{F}_t = \sum_{l=1}^{L} \mathbf{B}_l \mathbf{F}_{t-l} + \mathbf{v}_t \tag{7}$$

where $\mathbf{B}_l$ is a $q \times q$ matrix of autoregressive and cross-regressive coefficients, $\mathbf{F}_{t-l}$ is a vector of factor score l occasions prior to occasion t and $\mathbf{v}_t$ is a random shock vector (or innovation vector) following a multivariate normal distribution with mean zeros and $q \times q$ covariance matrix $\mathbf{D}$ (Nesselroade et al., 2002; Zhang et al., 2008). In the DAFS model, $\mathbf{\Lambda}$, $\mathbf{B}_l$, $\mathbf{Q}$ and $\mathbf{D}$ are invariant over time.

Data following the DAFS model can be simulated using the `simDFM` function of the *EGAnet* package (version 1.2.0; Golino & Christensen, 2019).

Results

Table 1
Accuracy per conditions tested

N	TimeP	NFAC	NVAR	NFAC2	NVAR2	Autoreg	Accuracy	Lower 95% C.I	Upper 95% C.I
10	50	2	4	4	2	0.1	75.00	72.31	77.69
10	50	2	4	4	2	0.4	91.30	89.55	93.05
10	50	2	4	4	2	0.8	98.70	98.00	99.40
10	50	2	6	3	4	0.1	100.00	100.00	100.00
10	50	2	6	3	4	0.4	100.00	100.00	100.00
10	50	2	6	3	4	0.8	100.00	100.00	100.00
10	50	3	4	2	6	0.1	74.40	71.69	77.11
10	50	3	4	2	6	0.4	68.20	65.31	71.09
10	50	3	4	2	6	0.8	68.90	66.03	71.77
10	50	3	6	2	9	0.1	78.50	75.95	81.05
10	50	3	6	2	9	0.4	90.30	88.46	92.14
10	50	3	6	2	9	0.8	98.80	98.12	99.48
10	100	2	4	4	2	0.1	70.00	67.15	72.85
10	100	2	4	4	2	0.4	93.50	91.97	95.03
10	100	2	4	4	2	0.8	99.70	99.36	100.04
10	100	2	6	3	4	0.1	100.00	100.00	100.00
10	100	2	6	3	4	0.4	100.00	100.00	100.00
10	100	2	6	3	4	0.8	100.00	100.00	100.00
10	100	3	4	2	6	0.1	83.60	81.30	85.90
10	100	3	4	2	6	0.4	83.20	80.88	85.52
10	100	3	4	2	6	0.8	83.78	81.49	86.07
10	100	3	6	2	9	0.1	82.40	80.04	84.76
10	100	3	6	2	9	0.4	96.90	95.82	97.98
10	100	3	6	2	9	0.8	99.90	99.70	100.10
50	50	2	4	4	2	0.1	66.10	63.16	69.04
50	50	2	4	4	2	0.4	97.90	97.01	98.79

Table 1
Accuracy per conditions tested (continued)

N	TimeP	NFAC	NVAR	NFAC2	NVAR2	Autoreg	Accuracy	Lower 95% C.I	Upper 95% C.I
50	50	2	4	4	2	0.8	99.90	99.70	100.10
50	50	2	6	3	4	0.1	100.00	100.00	100.00
50	50	2	6	3	4	0.4	100.00	100.00	100.00
50	50	2	6	3	4	0.8	100.00	100.00	100.00
50	50	3	4	2	6	0.1	95.90	94.67	97.13
50	50	3	4	2	6	0.4	95.30	93.99	96.61
50	50	3	4	2	6	0.8	94.50	93.08	95.92
50	50	3	6	2	9	0.1	86.60	84.49	88.71
50	50	3	6	2	9	0.4	99.80	99.52	100.08
50	50	3	6	2	9	0.8	100.00	100.00	100.00
50	100	2	4	4	2	0.1	66.80	63.88	69.72
50	100	2	4	4	2	0.4	99.20	98.65	99.75
50	100	2	4	4	2	0.8	100.00	100.00	100.00
50	100	2	6	3	4	0.1	100.00	100.00	100.00
50	100	2	6	3	4	0.4	100.00	100.00	100.00
50	100	2	6	3	4	0.8	100.00	100.00	100.00
50	100	3	4	2	6	0.1	99.80	99.52	100.08
50	100	3	4	2	6	0.4	99.90	99.70	100.10
50	100	3	4	2	6	0.8	99.50	99.06	99.94
50	100	3	6	2	9	0.1	96.60	95.47	97.73
50	100	3	6	2	9	0.4	100.00	100.00	100.00
50	100	3	6	2	9	0.8	100.00	100.00	100.00
100	50	2	4	4	2	0.1	69.80	66.95	72.65
100	50	2	4	4	2	0.4	99.40	98.92	99.88
100	50	2	4	4	2	0.8	100.00	100.00	100.00
100	50	2	6	3	4	0.1	100.00	100.00	100.00
100	50	2	6	3	4	0.4	100.00	100.00	100.00

Table 1
Accuracy per conditions tested (continued)

N	TimeP	NFAC	NVAR	NFAC2	NVAR2	Autoreg	Accuracy	Lower 95% C.I	Upper 95% C.I
100	50	2	6	3	4	0.8	100.00	100.00	100.00
100	50	3	4	2	6	0.1	100.00	100.00	100.00
100	50	3	4	2	6	0.4	99.90	99.70	100.10
100	50	3	4	2	6	0.8	100.00	100.00	100.00
100	50	3	6	2	9	0.1	95.80	94.55	97.05
100	50	3	6	2	9	0.4	100.00	100.00	100.00
100	50	3	6	2	9	0.8	100.00	100.00	100.00
100	100	2	4	4	2	0.1	72.75	69.98	75.51
100	100	2	4	4	2	0.4	99.90	99.70	100.10
100	100	2	4	4	2	0.8	100.00	100.00	100.00
100	100	2	6	3	4	0.1	100.00	100.00	100.00
100	100	2	6	3	4	0.4	100.00	100.00	100.00
100	100	2	6	3	4	0.8	100.00	100.00	100.00
100	100	3	4	2	6	0.1	100.00	100.00	100.00
100	100	3	4	2	6	0.4	100.00	100.00	100.00
100	100	3	4	2	6	0.8	100.00	100.00	100.00
100	100	3	6	2	9	0.1	99.60	99.21	99.99
100	100	3	6	2	9	0.4	100.00	100.00	100.00
100	100	3	6	2	9	0.8	100.00	100.00	100.00

Table 1 and 4 show the mean accuracy per condition tested and the 95% confidence interval of the mean. Mean accuracy is defined as the average (across 500 simulated datasets per condition) number of times the EII was larger for the *NotEq* condition than for the *Eq* condition (in a scale from 0 to 100%). The grand mean accuracy was very high (Mean = 94.07%, SD = 23.62). The accuracy increases with an increase in the sample size, number of variables per factor, and as a function of the autoregressive parameter. When the number of

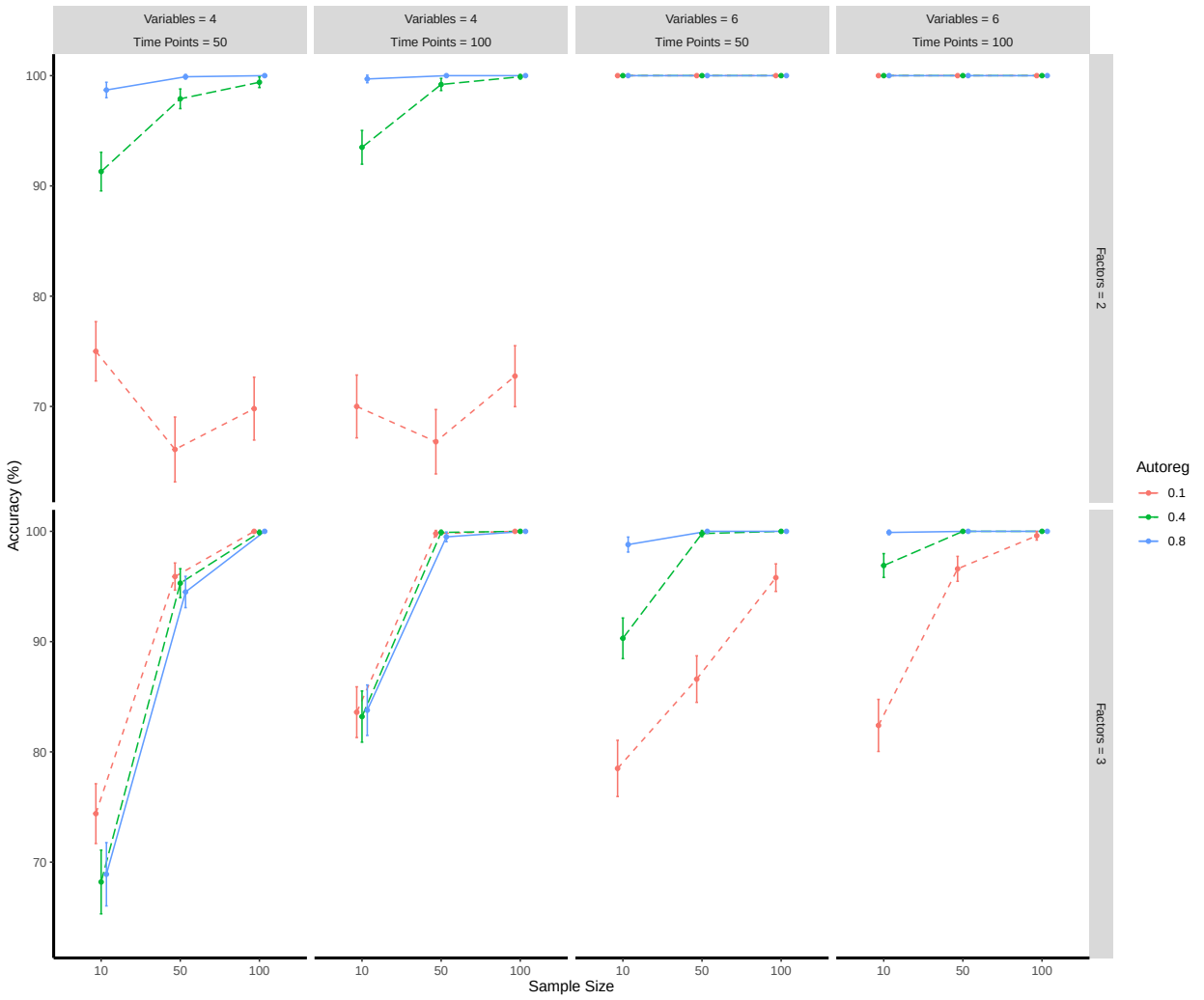

Figure 4. Mean accuracy per condition tested

factors is two, the EII presented a moderate accuracy for an autoregressive parameter of .1, with only four variables per factor. As the number of variables per factor increased to six, the accuracy was perfect. This is likely due to the configuration of the second group of individuals in the *NotEq* condition. When the number of factors is two in the first group with four variables per factor, the second group had four factors, each with two variables only (total number of variables = 8). Increasing the number of variables to six in the first group (with two main factors) means that the second group could have three factors with four variables or four factors with three variables each (total number of variables = 12). So, the six variables condition had better identified factors than the four variables condition

when the number of factors equals two for the second group of the *NotEq* condition. For the three factors condition, sample size was more impactful when the number of variables is set to four, and the autoregressive parameter was more impactful in the conditions with six variables per factor.

Figure 5 shows the mean EII value per sample size, number of factors, number of variables per factor, number of time points, and autoregressive parameter. Overall, the mean EII of the *NotEq* condition is much higher than the mean of the *Eq* condition, showing that the metric can reliably differentiate when the individuals have a similar structure compared to when individuals have a different structure. This is an important evidence of the reliability of the EII to capture the amount of information lost by representing a set of measures as a single population (or interindividual) structure instead of representing the measures as multiple individual structures.

Figure 6 shows the distribution of the EII value per number of factors and number of variables per factor (of the first group of individuals). Finally, Figure 7 shows the distribution of the EII value per number of variables per factor in the first group (rows) and in the second group of individuals (columns).

A bottom-up approach for finding generalizations

If the system is nonergodic, what then? From Figure 3, it's reasonable to think that elements of the network structure estimated for each individual can be used to search for similarities that could help uncover sub-groups of people or generalizable characteristics across individuals. Said differently, uncovering sub-groups could reveal ergodic systems underlying the overall system. Since Campbell and Stanley (1963) raised concerns regarding external validity in psychological research, generalizability hasn't received nearly enough attention (Yarkoni, 2022).

The emphasis on generalizability has both weaknesses and strengths in the directions

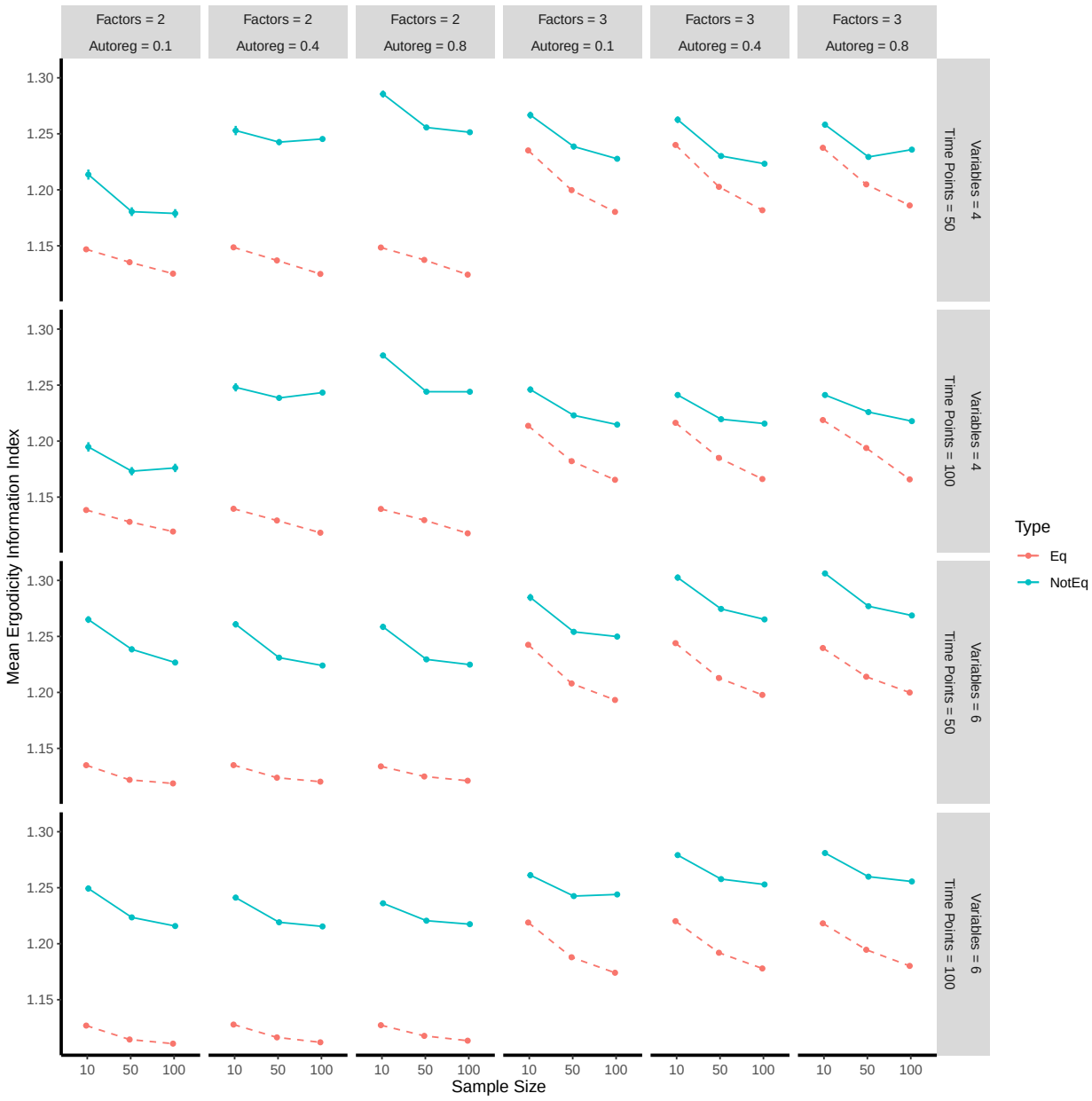

Figure 5. Mean ergodicity information index (and 95 percent confidence interval of the mean) per condition tested

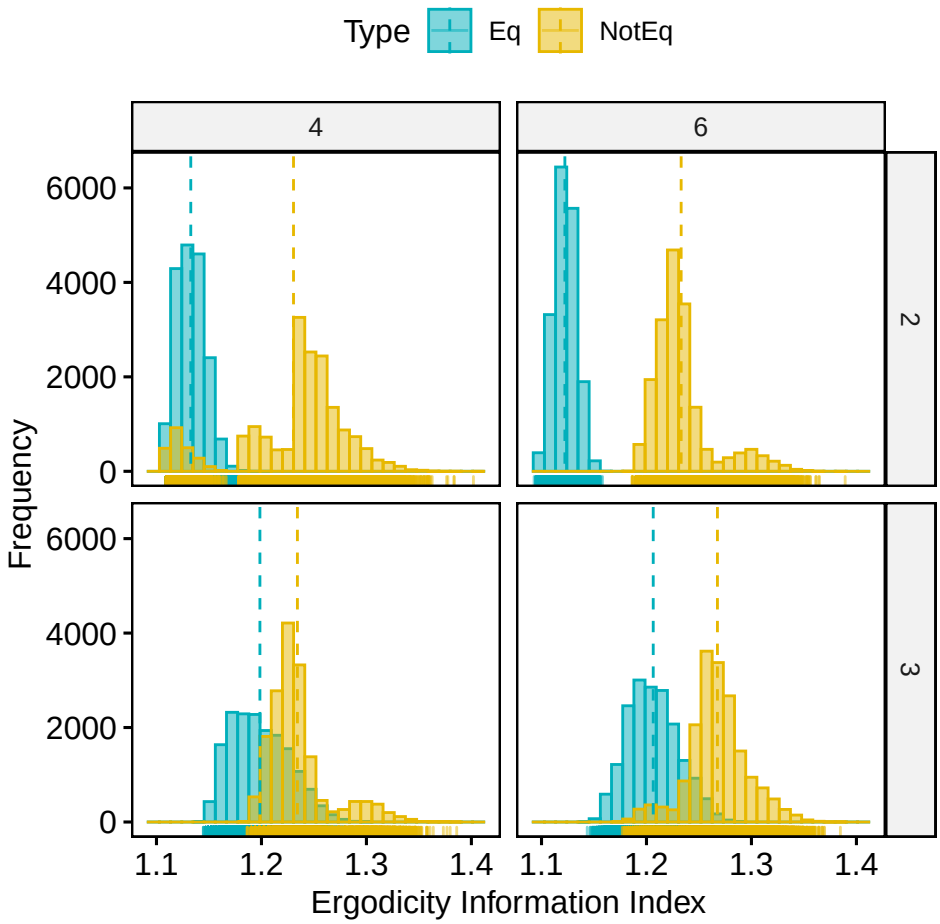

Figure 6. Distribution of the mean ergodicity fit index per number of factors (rows) and number of variables per factor (columns)

in which it has helped push research design and data analysis. Among the negatives, an emphasis on drawing large, representative samples of participants has, to our way of thinking, been detrimental to the study of complex psychological processes. The primary reason for this assertion is that to the extent that processes are constructs, they may well manifest themselves somewhat differently from individual to individual (Molenaar & Nesselroade, 2012) making it meaningless to build an accurate representation of the process with observable measures. Here, heterogeneity is not helpful. If one is trying to determine the average number of bedrooms in single family dwellings in the U.S., then a truly representative sampling of single family dwellings is highly desirable. But if one is trying to determine the nature of the onset and progression of depression, a representative sample that

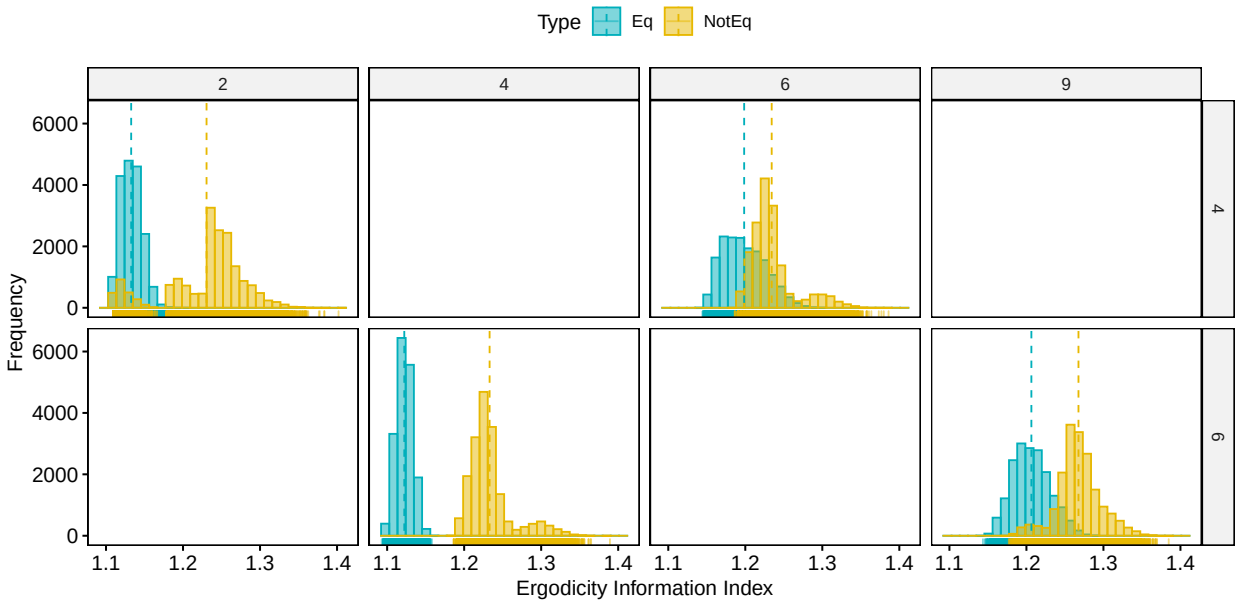

Figure 7. Distribution of the mean ergodicity fit index per number of variables per factor of the first group (rows) and number of variables per factor in the second group (columns)

includes a variety of paths of onset and progression is not helpful and may aggregate over paths to a point where the representation does not match any person in the sample.

In arguing for a "bottom-up" approach to the matter of generalizability, Nesselroade and Molenaar (2016) argued: "It seems far more appropriate to apply tools that emphasize first understanding individuals well and then identifying similarities across persons, thus accruing generalizability gradually than initially fitting models to heterogeneous samples in order to claim generalizability. Large, diverse samples of individuals may put a gleam in a demographer's eye but a wide array of observed differences, however universal the underlying mechanisms may be, can blind the behavioral scientist to the identification of general processes (p.15)." Following this line of reasoning, we propose an approach that takes into consideration the spectral properties of the individual network structures in the search for sub-groups or generalizable characteristics using an information-theoretic metric to compare the similarity between two structures.

Spectral properties of a network provide insights into not only the topological features

(i.e., connectivity between nodes) but also the community structure (Chauhan, Girvan, & Ott, 2009) and temporal dynamics (Almendral & Dıaz-Guilera, 2007). Recent work has taken advantage of these properties to determine whether individual networks in a multiplex network can be aggregated into groups (De Domenico, Nicosia, Arenas, & Latora, 2015). De Domenico et al. (2015) proposed a multiplex network reduction approach by computing Von Neumman entropy of two networks and computing their Jensen-Shannon Distance (see Golino, Moulder, et al., 2020 for other applications of Von Neumann entropy in community detection). Von Neumann entropy of a network can be computed as follows:

$$h_A = -\mathrm{Tr}[\mathcal{L}_G \log_2 \mathcal{L}_G],$$

where $\mathcal{L}_G = c \times (D - A)$ is the combinatorial Laplacian rescaled by c or one over the sum of the weights in the network. D is a matrix with the strength of each node (i.e., sum of each node's connections) on its respective diagonal and A is the network. $\mathcal{L}_G$ is a density matrix that is then used to compute Von Neumman entropy:

$$h_A = -\sum_{i=1}^{N} \lambda_i \log_2(\lambda_i),$$

where λ are the eigenvalues of $\mathcal{L}_G$. Using Von Neumann entropy of the network, we can compute the Jensen-Shannon Divergence between two networks, which is a symmetric measure of dissimilarity related to Kullback-Leibler Divergence (De Domenico et al., 2015):

$$\mathcal{D}_{JS}(\rho||\sigma) = h(\mu) - \frac{1}{2}[h(\rho) + h(\sigma)],$$

where ρ and σ are $\mathcal{L}_G$ of each network being compared and $\mu = \frac{1}{2}(\rho + \sigma)$. Taking the square root of $\mathcal{D}_{JS}$ produces a $[0, 1]$ bound metric often referred to as Jensen-Shannon Distance. Subtracting 1 from the Jensen-Shannon Distance produces a similarity metric rather than distance metric.

De Domenico et al. (2015) used an entropy-based quality function and Ward's method for agglomerative hierarchical clustering using the Jensen-Shannon Distance to determine whether individual layers (i.e., networks) in a multiplex network can be aggregated. Our approach follows allow similar lines: We use complete-linkage agglomerative hierarchical clustering on the Jensen-Shannon Distance. The trees are then cut through all possible cuts, obtaining as many sets of clusters as there are networks. Afterward, a similarity matrix (1 - Jensen-Shannon Distance) is obtained and used to compute modularity. Modularity is an objective function used in many community detection algorithms in network science (**newman2006modularity?**). The modularity metric quantifies the extent to which within-cluster similarity is maximized and between-cluster similarity is minimized. The clusters that maximize modularity are taken to be the groups of similar networks.

Data from 30 individuals, with 100 measurement points each and 12 variables, were simulated using the direct autoregressive factor score model (**DAFS**; Engle & Watson, 1981; Nesselroade et al., 2002). The individuals differed only in the number of underlying factors (2, 3, or 4), forming three groups. Using DynEGA to estimate the intraindividual networks and computing the Jensen-Shannon Distance for every pair of networks, enabled the identification of three clusters of people (bottom left of Figure 8). On the top left of Figure 8, the hierarchical clustering dendogram with three groups correctly identified. On the top right of Figure 8, the three different groups' network and community structures are represented separately. On the bottom right of Figure 8, the population structure estimated using DynEGA is shown for each group of individuals.

Testing EII and Determining Groups in Applied Examples

Bootsrtap EII Test

The EII provides a relative metric for the information lost when representing the sample as a between-person, population structure relative to within-person, individual

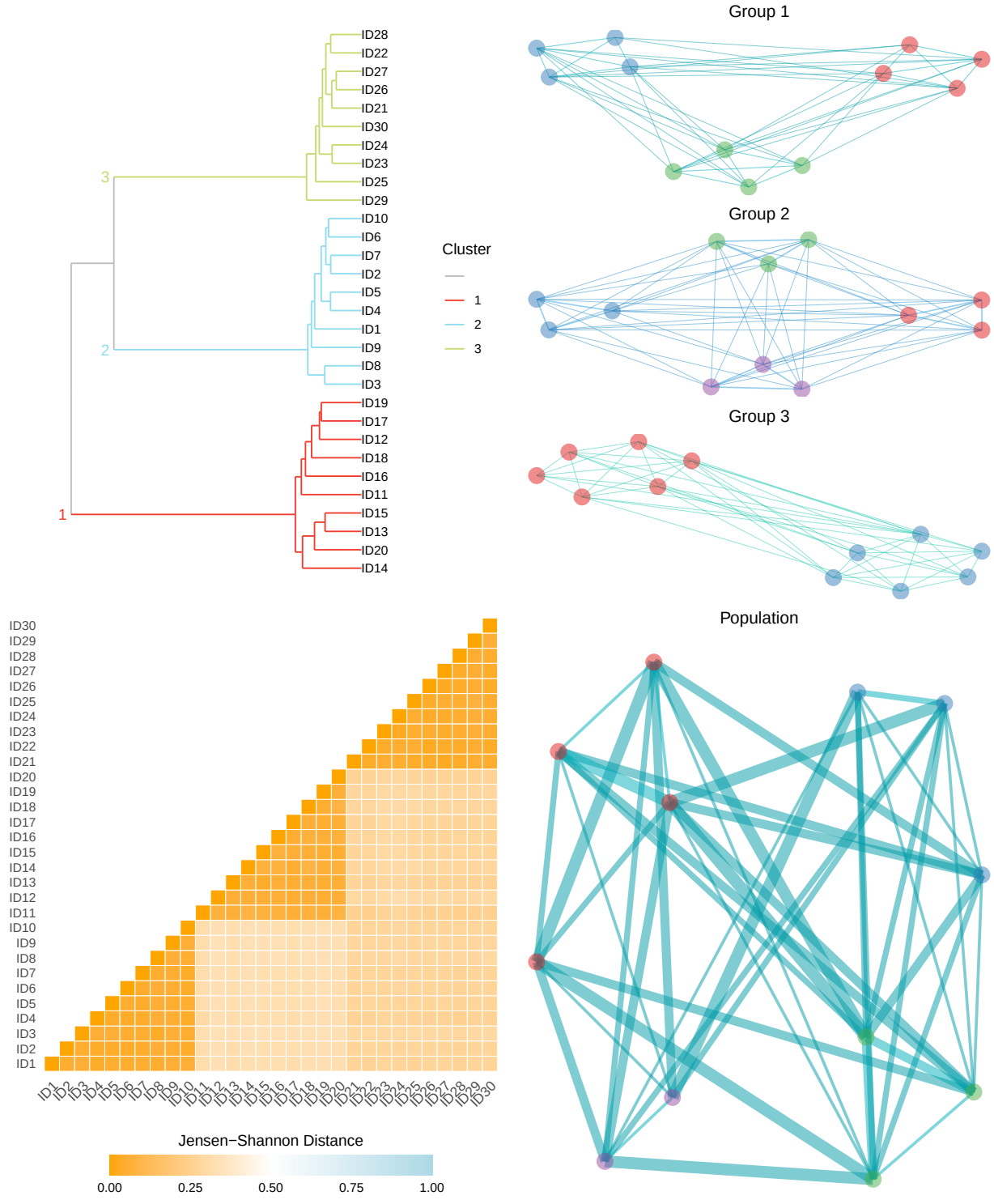
ID28
ID22
ID27
ID26
ID21
ID30
ID24
ID23
ID25
ID29
ID10
ID6
ID7
ID2
ID5
ID4
ID1
ID9
ID8
ID3
ID19
ID17
ID12
ID18
ID16
ID11
ID15
ID13
ID20
ID14
1
2
3
Cluster
1
2
3
Group 1
Group 2
Group 3
Population
Jensen−Shannon Distance
0.00
0.25
0.50
0.75
1.00

Figure 8

structures. Determining whether the amount of information lost is substantial requires understanding the information loss relative to when within-person structures match the between-person structure. One approach would be to generate random multivariate normal samples from the population structure's covariance matrix with the same number of participants and their corresponding time series to create a sampling distribution of EII to compare with the empirical EII. This approach, however, is computationally expensive (e.g., 100 bootstraps $\times$ 50 participants $\times$ 20 variables $\times$ 10 time points).

A simpler approach is to acquire a distribution of EII when individuals are known to be similar to the population structure. To do so, we rewire the population network structure to generate variants that are close but not identical to the population structure. This approach takes the population network structure and completely rewires an edge, randomly reassigning an edge connecting two nodes to two other nodes. To obtain structures with different levels of deviations away from the population structure, we randomly sample a proportion of edges to be rewired from a uniform distribution with a minimum of 0.20 (20% of edges) to 0.40 (40% of edges). In addition, we add random noise to all edges in the network randomly sampling from a uniform distribution ranging from -0.10 to 0.10. This process simulates small to moderate perturbations to the population structure to obtain structures that might be observed by individuals within a population.

We generate as many population variants as their are participants in the empirical sample. With the new sample containing the population network variants, we compute EII with the original population network as the population network and the population network variants as the individuals. We repeat this process for X iterations (e.g., 100). This approach creates a sampling distribution of EII that would expected when the individuals in the population have slight deviations on the population structure. When the empirical EII is greater than 97.5% of the samples (two-tailed $p = 0.05$), then there is significant information lost when representing the sample as an aggregate, population network. In all other cases, including when the empirical EII is less than 97.5% of the sample, the ergodic property holds

such that non-substantial information is lost when representing the sample as an aggregate, population network. Applying this approach to the data generated in Figure **??**, the data are nonergodic when combined into an aggregate population but ergodic when separated by cluster.

Single Cluster Test

If the bootstrap EII test is significant, then the next step is to determine whether there are clusters (or groupings) that can be identified in the data. In some cases, there may be mixtures of network structures that form groups of people that have similar structures. In other cases, each person may be unique and there are no clusters. In the latter case, a challenge arises: the JSD values between each network structure will be relatively equidistant. Because of the similar distances between networks, many clustering algorithms will suggest that one cluster exists. If the bootstrap EII test is significant, this result is paradoxical: The data do not possess the ergodic property yet they form a single cluster. While the bootstrap EII could be leveraged to interpret the result as the single cluster reflects the lack of clustering, the evidence is unsatisfying.

We derive an additional test to determine whether a single cluster represents a single homogeneous cluster or if a single cluster represents heterogeneity such that each individual represents their own cluster (i.e., singleton clusters). This additional procedure is only applied when the number of clusters is determined to equal one cluster. The procedure begins by generating a random network with the same number of nodes and edges as each individual network in the empirical sample. Next, JSD is computed between every pair of random networks (as is done to the empirical networks). The JSD values of the empirical and random networks are then compared using a paired samples t-test, comparing the corresponding JSD values between the corresponding empirical (e.g., empirical network j and empirical network j) and random networks (e.g., random network i and random network j).

If the empirical network JSD values are, on average, greater than or equal to the random network JSD values, then the empirical network similarities are no different than the random network similarities suggesting that the single cluster is not meaningful and no meaningful clusters exist. If the empirical JSD values are, on average, significantly less than the random network JSD values (based on a large effect adjusted for sample size; Pérez & Pericchi, 2014), then the empirical network similarities are substantially less than the random networks similarities suggesting that the single cluster is meaningful.

We performed a small simulation to determine the effectiveness of this single cluster test. One hundred heterogeneous samples of thirty cases with fifty time points were generated using the **DAFS** method. Each case in each sample was randomly generated with 36 variables that loaded (between 0.70 and 1.00) onto 2, 3, 4, 6, or 9 factors with moderate to large error (0.20-0.50). Similarly, one hundred homogeneous samples were equivalent generated except that, rather than each case being generated by a different factor structure, all cases for each sample were generated from the same factor structure with no to moderate error (0.00-0.30). For the heterogeneous samples, an iteration was thrown out and replaced if there were two or more clusters identified. Given the conditions randomly sample from only five potential factor structures (with the same number of items per factor), then its possible that these clusters truly exist in the data and therefore are not representative of a heterogeneous sample. For the homogeneous samples, an iteration was thrown out and replaced if the bootstrap EII test was significant. The accuracy for correctly identifying when heterogeneous samples were heterogeneous (i.e., truly singleton clusters) was 100% and the accuracy for correctly identifying when homogeneous samples were homogeneous (i.e., truly one cluster) was 97%.

With this single cluster test, all evidence can be used together. If the bootstrap EII test is significant and the single cluster test suggests singleton clusters, then the data are not ergodic and they possess no meaningful groupings. If the bootstrap EII test is non-significant and the single cluster test suggests a single cluster, then the data are ergodic and represent a

single, homogeneous sample. What if the bootstrap EII test is significant and the single cluster test suggests a single cluster? While we expect this last circumstance to be rare, it's possible. Our stance is that the default position should be to assume that the data do *not* possess the ergodic property—that is, we believe evidence must be accumulated for ergodicity and a homogeneous sample. In null hypothesis testing terms, the null hypothesis should be that the data are not ergodic and do not represent a single, homogeneous sample and the alternative hypothesis requiring significant evidence to the contrary.

Empirical Examples

To demonstrate the bootstrap EII and single cluster tests, we evaluated two empirical data examples that are commonly represented with an aggregate structure: personality and brain networks. For personality, we used an empirical example taken from an intensive longitudinal experience sampling study examining Big Five personality measured by the Big Five Inventory-2 (Beck & Jackson, 2022; Soto & John, 2017). There were 199 participants who completed between 1 and 158 time points. To ensure optimal data quality, we only included participants who completed at least 20 time points and had network densities of at least 0.15 (i.e., at least 15% of all possible connections present). These criteria narrowed the final sample to 122 participants. For the brain data, we used resting state neuroimaging data taken from a study examining creativity (Beaty et al., 2018), which used the 268-node Shen brain atlas (Shen, Tokoglu, Papademetris, & Constable, 2013). We analyzed the time series data of 176 participants. The resting state scan was 5 minutes in duration resulting in 150 time points (300 seconds with TR of 2 seconds).

Starting with personality, the bootstrap EII test was significant ($EII = 1.417$, $p = 0.02$, $EII_{bootstrap} = [1.405, 1.408]$) suggesting that the BFI-2 data were nonergodic. Following up the bootstrap EII test, the information clustering was applied and the single cluster test was performed. The single cluster test suggested that the empirical networks had significantly greater JSD values than the random networks

($M_{difference} = 0.11, t(7380) = 66.85, p < .001, d = 1.08$) meaning that the single cluster detected was not meaningful and each individual in the sample is unique.

Like the personality data, the brain networks had a significant bootstrap EII test ($EII = 6.237$, $p = 0.02$, $EII_{bootstrap} = [5.855, 5.943]$) and significantly larger JSD values than random networks, $M_{difference} = 0.02, t(15399) = 159.28, p < .001, d = 1.78$. These results, in line with the personality data, suggest that (resting state) brain networks are not ergodic and no meaningful groups can be formed (i.e., each individual is unique). In sum, the personality and (resting state) brain networks do not possess ergodicity and therefore lose information when data are aggregated into a single population.

Discussion

For well over a century the study of variation has been the backbone of psychologists' efforts to understand behavior and behavior change. Whether created via experimental manipulation or measured as it exists in nature (Cronbach, 1957; e.g., Cronbach, 1975), variation is the ore that has been dug up, assayed, weighed, and otherwise analyzed by behavioral prospectors hoping to strike it rich. Between-person or interindividual variation has been and is strongly favored in psychological research. Often, researchers infer within-person or idiographic processes based on between-person variation including many prominent theories. The reliance on interindividual variation has led to a field that appears to implicitly assume ergodicity is a fundamental property of psychological processes. Despite this ubiquitous assumption, many researchers have chosen to focus on how the behavior of an individual could vary from one measurement occasion to the next (Beck & Jackson, 2022; Fisher et al., 2018).

Intraindividual variability is no longer studied by a curious few. It has grown into a prominent psychological method over the last few decades, to the point that it warrants the undivided attention of both methodologists and substantive theorists. This article takes

steps toward more rigorous testing about whether researchers can safely ignore the different patterns of intraindividual variability in order to construct more general interindividual representations. We leveraged tools from dynamical systems, network science, and information theory to develop tests to the determine cost of aggregation. DynEGA paired with GLLA allows researchers to study how variables change together over time and the general structure of their relations (Golino et al., 2022). The EII provides researchers with a relative metric and bootstrap test to determine how much information is lost when aggregating individuals into a single population structure and whether that amount is significant. If significant information is lost, then the information clustering method can be applied and, if necessary, a single cluster test can evaluate whether meaningful clusters exist in the data.

We propose that EII quantifies the extent to which *super-weak ergodicity* holds in the system. Super-weak ergodicity suggests that the individual structures of a system should reflect, within reasonable error, the aggregate structure of the system. This level of ergodicity is a minimum requirement of a system to be represented as an aggregate. Systems that do not possess this property should not be aggregated because significant information is lost—such that a mere fraction of the system can be expected to reflect the aggregate system. In its present state, psychological processes are unlikely to be ergodic (Molenaar, 2004). Therefore, the measurement of ergodicity must be pursued in perpetuity. Our position, as well as the position of many others (e.g., Fisher et al., 2018; Molenaar, 2004), is that psychological processes should be considered nonergodic until they are repeatedly demonstrated otherwise.

Our empirical examples take two commonly aggregated psychological phenomena, personality and (resting state) brain activity, and examined the extent to which they lose information when aggregated into a single population network. We show, by no slim margin, that personality and brain networks are nonergodic. We further show that the distance between each individual's structure is more than would be expected if it were generated by

an equivalent random structure. For personality, these result question the extent to which the Big Five generalize to individuals (Borkenau & Ostendorf, 1998). For brain networks, these results add to the growing body of evidence that dynamic brain networks reveal greater idiosyncrasies between people (Hutchison et al., 2013; Lurie et al., 2020). Together, these results suggest that each individual person shows distinct processes that are lost when modeled as a single aggregate structure. In short, "no two people are alike" (@ Nesselroade & Molenaar, 2016).

What are the consequences of our findings? In the best case, we show that the personality and brain network samples examined in our study are nonergodic and are perhaps due to sampling variability. There are many underlying factors that make this interpretation plausible such as the specific personality scale used (BFI-2) or the brain network task (or lack of task) or the demographics of the samples (i.e., predominantly college students). Our findings, however, are not the first to find that individual people have resting state brain signatures that make them uniquely identifiable (e.g., Finn et al., 2015). In the worst case, we show that personality and brain activity are nonergodic and that research examining their aggregates, the vast majority of their literatures, rest on a faulty assumption. If the latter is true, then our findings challenge the validity of the conclusions drawn by decades of interindividual research in personality and neuroscience. Our hope is that the truth is somewhere in between.

Psychology aims to understand the thoughts, emotions, and behavior of the person and people. Despite repeated calls and manifestos, the study of people continues to dominate psychology. With modern technology, everyday thoughts, emotions, and behaviors of the person have never been more accessible. As psychologists begin to emphasize intraindividual processes over interindividual predictions, statistical models that answer questions about the dynamics of systems are needed (Epskamp et al., 2018b; e.g., Gates & Molenaar, 2012; Moulder, Martynova, & Boker, 2021; Sterba & Bauer, 2010). The present work provides one statistical tool to understand whether intraindividual dynamics can be reasonably

aggregated into a single interindividual structure and another tool to determine possible sub-aggregations when they cannot. Together, these tools allow researchers to establish generalizability starting with the person rather than searching for it across people (Nesselroade & Molenaar, 2016).

References

Almendral, J. A., & Dıaz-Guilera, A. (2007). Dynamical and spectral properties of complex networks. *New Journal of Physics*, *9*(6), 187. https://doi.org/10.1088/1367-2630/9/6/187

Anderson, H., T. W. & Rubin. (1958). Statistical inference in factor analysis. *Proceedings of the 3rd Berkeley Symposium on Mathematics, Statistics, and Probability*, *5*, 111–150.

Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q., ... Silvia, P. J. (2018). Robust prediction of individual creative ability from brain functional connectivity. *Proceedings of the National Academy of Sciences*, *115*, 1087–1092. https://doi.org/10.1073/pnas.1713532115

Beck, E. D., & Jackson, J. J. (2022). Idiographic personality coherence: A quasi experimental longitudinal ESM study. *European Journal of Personality*, *36*(3), 391–412. https://doi.org/10.1177/08902070211017746

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change*. Madison, WI: University of Wisconsin Press.

Boker, S. M. (2018). *Longitudinal multivariate psychology* (E. Ferrer, S. M. Boker, & K. J. Grimm, Eds.). Routledge.

Boker, S. M., Deboek, P. R., Edler, C., & Keel, P. (2010). Generalized local linear approximation of derivatives from time series. In S. M. Chow, E. Ferrer, & F. Hsieh (Eds.), *The notre dame series on quantitative methodology. Statistical methods for modeling human dynamics: An interdisciplinary dialogue* (pp. 161–178). Routledge/Taylor & Francis Group.

Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor model to describe intraindividual variations over time? *Journal of Research in Personality*, *32*(2), 202–221. https://doi.org/10.1006/jrpe.1997.2206

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi–experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally.

Cattell, R. B. (1965). *Studies in psychology* (C. Banks & P. L. Broadhurst, Eds.). University of London Press London.

Cattell, R., Cattell, A., & Rhymer, R. M. (1947). P-technique demonstrated in determining psychophysiological source traits in a normal individual. *Psychometrika*, *12*(4), 267–288.

Chauhan, S., Girvan, M., & Ott, E. (2009). Spectral properties of networks with community structure. *Physical Review E*, *80*(5), 056114. https://doi.org/10.1103/PhysRevE.80.056114

Chen, J., & Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*(3), 759–771. Retrieved from https://www.jstor.org/stable/20441500

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*(2), 116.

De Domenico, M., Nicosia, V., Arenas, A., & Latora, V. (2015). Structural reducibility of multilayer networks. *Nature Communications*, *6*(1), 1–9. https://doi.org/10.1038/ncomms7864

Deboeck, P. R., Montpetit, M. A., Bergeman, C., & Boker, S. M. (2009). Using derivative estimates to describe intraindividual variability at multiple time scales. *Psychological Methods*, *14*(4), 367–386. https://doi.org/http://dx.doi.org/10.1037/a0016622

Diehl, M., Hooker, K., & Sliwinski, M. J. (2014). *Handbook of intraindividual variability across the life span*. East Sussex, BN: Routledge.

Engle, R., & Watson, M. (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, *76*(376), 774–781.

Epskamp, M., S. (2018). Network psychometrics. In B. Irwing Paul (Ed.), *The wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (pp. 953–986). John Wiley & Sons Ltd. https://doi.org/10.1002/9781118489772.ch30

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. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*(4), 617–634. https://doi.org/10.1037/met0000167

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

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018b). The gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018a). The gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*(4), 453–480.

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., … Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671. https://doi.org/10.1038/nn.4135

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual

generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106–E6115.

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of Abnormal Psychology*, *126*(8), 1044.

Foygel, R., & Drton, M. (2010). Extended bayesian information criteria for gaussian graphical models. *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 1*, *1*, 604–612. Vancouver, Canada.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441. https://doi.org/10.1093/biostatistics/kxm045

Gates, K. M., & Molenaar, P. C. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, *63*(1), 310–319. https://doi.org/10.1016/j.neuroimage.2012.06.026

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS One*, *12*(6), e0174035. https://doi.org/10.1371/journal.pone.0174035

Golino, H., & Christensen, A. P. (2019). *EGAnet: Exploratory graph analysis: A framework for estimating the number of dimensions in multivariate data using network psychometrics*. Retrieved from https://CRAN.R-project.org/package=EGAnet

Golino, H., Christensen, A. P., Moulder, R., Kim, S., & Boker, S. M. (2022). Modeling latent topics in social media using Dynamic Exploratory Graph Analysis: The case of the right-wing and left-wing trolls in the 2016 US elections. *Psychometrika*, *87*(1), 156–187.

Golino, H., Moulder, R., Shi, D., Christensen, A., Garrido, L., Neto, M., ... Boker, S. (2020). Entropy fit indices: New fit measures for assessing the structure and

826 dimensionality of multiple latent variables. *Multivariate Behavioral Research*.
827 https://doi.org/10.1080/00273171.2020.1779642
828 Golino, H., Shi, D., Garrido, L. E., Christensen, A. P., Nieto, M. D., Sadana, R., ...
829 Martinez-Molina, A. (2020). Investigating the performance of exploratory graph
830 analysis and traditional techniques to identify the number of latent factors: A
831 simulation and tutorial. *Psychological Methods*, *25*(3), 292–230.
832 https://doi.org/10.1037/met0000255
833 Gomes, C. M. A., & Golino, H. F. (2015). Factor retention in the intra-individual
834 approach: Proposition of a triangulation strategy. *Avaliaçao Psicologica:*
835 *Interamerican Journal of Psychological Assessment*, *14*(2), 2.
836 Guttman, L. (1953). Image theory for the structure of quantitative variates.
837 *Psychometrika*, *18*(4), 277–296.
838 Hultsch, D. F., Strauss, E., Hunter, M. A., & MacDonald, S. W. S. (2008).
839 Intraindividual variability, cognition, and aging. In F. I. M. Craik & T. A.
840 Salthouse (Eds.), *The handbook of aging and cognition* (pp. 491–556). Psychology
841 Press.
842 Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D.,
843 Corbetta, M., ... others. (2013). Dynamic functional connectivity: Promise,
844 issues, and interpretations. *Neuroimage*, *80*, 360–378.
845 https://doi.org/10.1016/j.neuroimage.2013.05.079
846 Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*,
847 *106*(4), 620.
848 Kolmogorov, A. N. (1968). Three approaches to the quantitative definition of
849 information. *International Journal of Computer Mathematics*, *2*(1-4), 157–168.
850 Lamiell, J. T. (1998). Nomothetic and idiographic: Contrasting windelband's
851 understanding with contemporary usage. *Theory & Psychology*, *8*(1), 23–38.
852 https://doi.org/10.1177/0959354398081002

Lauritzen, S. L. (1996). *Graphical models* (Vol. 17). Oxford: Clarendon Press.

Liu, S., Zhou, Y., Palumbo, R., & Wang, J.-L. (2016). Dynamical correlation: A new method for quantifying synchrony with multivariate intensive longitudinal data. *Psychological Methods*, *21*(3), 291.

Lurie, D. J., Kessler, D., Bassett, D. S., Betzel, R. F., Breakspear, M., Kheilholz, S., ... others. (2020). Questions and controversies in the study of time-varying functional connectivity in resting fMRI. *Network Neuroscience*, *4*(1), 30–69. https://doi.org/10.1162/netn_a_00116

Martin-Löf, P. (1966). The definition of random sequences. *Information and Control*, *9*(6), 602–619.

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Molenaar, P. C., Huizenga, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology* (pp. 339–360). Springer.

Molenaar, P. C., & Nesselroade, J. R. (2012). Merging the idiographic filter with dynamic factor analysis to model process. *Applied Developmental Science*, *16*(4), 210–219. https://doi.org/10.1080/10888691.2012.722884

Morzy, M., Kajdanowicz, T., & Kazienko, P. (2017). On measuring the complexity of networks: Kolmogorov complexity versus entropy. *Complexity*, *2017*.

Moulder, R. G., Martynova, E., & Boker, S. M. (2021). Extracting nonlinear dynamics from psychological and behavioral time series through HAVOK analysis. *Multivariate Behavioral Research*, 1–25. https://doi.org/10.1080/00273171.2021.1994848

Nesselroade, J. R., & Ford, D. H. (1985). P-technique comes of age: Multivariate, replicated, single-subject designs for research on older adults. *Research on Aging*, *7*(1), 46–80. https://doi.org/10.1177/0164027585007001003

Nesselroade, J. R., McArdle, J. J., Aggen, S. H., & Meyers, J. M. (2002). Dynamic factor analysis models for representing process in multivariate time-series. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Multivariate applications book series. Modeling intraindividual variability with repeated measures data: Methods and applications* (pp. 235--265). Lawrence Erlbaum Associates Publishers.

Nesselroade, J. R., & Molenaar, P. C. (2016). Some behaviorial science measurement concerns and proposals. *Multivariate Behavioral Research*, *51*(2-3), 396–412.

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford University Press.

Oertzen, T. von, Schmiedek, F., & Voelkle, M. C. (2020). Ergodic subspace analysis. *Journal of Intelligence*, *8*(1), 3.

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

Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks. In Pi. Yolum, T. Güngör, F. Gürgen, & C. Özturan (Eds.), *Computer and information sciences - ISCIS 2005* (pp. 284–293). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11569596_31

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ram, N., Gerstorf, D., Lindenberger, U., & Smith, J. (2011). Developmental change and intraindividual variability: Relating cognitive aging to cognitive plasticity, cardiovascular lability, and emotional diversity. *Psychology and Aging*, *26*(2), 363.

Rosenstein, M. T., Collins, J. J., & De Luca, C. J. (1993). A practical method for

calculating largest lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena*, *65*(1-2), 117–134. https://doi.org/10.1016/0167-2789(93)90009-P

Santoro, A., & Nicosia, V. (2020). Algorithmic complexity of multiplex networks. *Physical Review X*, *10*(2), 021069.

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2020). Training working memory for 100 days: The COGITO study. In *Cognitive and working memory training: Perspectives from psychology, neuroscience, and human development* (pp. 40–57). Oxford University Press.

Schmiedek, F., Lövdén, M., Oertzen, T. von, & Lindenberger, U. (2020). Within-person structures of daily cognitive performance differ from between-person structures of cognitive abilities. *PeerJ*, *8*, e9290.

Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage*, *82*, 403–415. https://doi.org/10.1016/j.neuroimage.2013.05.081

Solomonoff, R. J. (1964). A formal theory of inductive inference. *Information and Control*, *7*(1), 1–22. https://doi.org/10.1016/S0019-9958(64)90223-2

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*(1), 117–143. https://doi.org/10.1037/pspp0000096

Sterba, S. K., & Bauer, D. J. (2010). Matching method with theory in person-oriented developmental psychopathology research. *Development and Psychopathology*, *22*(2), 239–254. https://doi.org/10.1017/S0954579410000015

Takens, F. (1981). Detecting strange attractors in turbulence. In *Lecture notes in mathematics (vol. 898)* (pp. 366–381). Springer. https://doi.org/10.1007/BFb0091924

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

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. https://doi.org/10.1007/BF02293557

Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research*, *49*(3), 193–213.

Whitney, H. (1936). Differentiable manifolds. *The Annals of Mathematics*, *37*(3), 645–680. https://doi.org/10.2307/1968482

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, *28*(3), 263–311. https://doi.org/10.1207/s15327906mbr2803_1

Williams, D. R., & Rast, P. (2020). Back to the basics: Rethinking partial correlation network methodology. *British Journal of Mathematical and Statistical Psychology*, *73*(2), 187–212. https://doi.org/10.1111/bmsp.12173

Williams, D. R., Rhemtulla, M., Wysocki, A. C., & Rast, P. (2019). On nonregularized estimation of psychological networks. *Multivariate Behavioral Research*, *54*(5), 719–750. https://doi.org/10.1080/00273171.2019.1575716

Wright, A. G., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychological Assessment*, *31*(12), 1467.

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1. https://doi.org/10.1017/S0140525X20001685

Zenil, H., Kiani, N. A., & Tegnér, J. (2018). A review of graph and network

961 complexity from an algorithmic information perspective. *Entropy*, *20*(8), 551.
962 Zhang, Z., Hamaker, E. L., & Nesselroade, J. R. (2008). Comparisons of four
963 methods for estimating a dynamic factor model. *Structural Equation Modeling: A*
964 *Multidisciplinary Journal*, *15*(3), 377–402.
965 https://doi.org/doi.org/10.1080/10705510802154281