# Temporal stability of idiographic psychological networks

Ricarda K. K. Proppert

Leiden University, The Netherlands

s1348981

Author note

Thesis project for the Research Master Clinical & Health Psychology, supervised by Eiko Fried, Ph.D., Leiden University. We thank Aidan Wright, Ph.D., Pittsburgh University, for sharing data for this reanalysis. The data and analysis script are available at https://github.com/RicardaP/thesis_repo and https://osf.io/gnw4s.

Abstract

Evidence-based mental health programs have long conceptualized mental disorders as interactions between thoughts, feelings, behaviours and external factors. Idiographic network models are a relatively novel way of estimating such intra-individual psychological processes. These methods are not without limitations, and concerns have been raised about the stability and accuracy of estimated networks. The extend to which idiographic networks are stable, or vary over time, is unknown. We explored temporal network stability from three angles, exploring variation within people, across different stability metrics, and across people. We reanalysed daily symptom records of people with personality disorders. We fit graphical Vector Autoregressive models separately for the first and second 50 days of consecutive measurements. Contemporaneous but not temporal idiographic networks appeared to be relatively stable within people. The assessment of stability varied considerably across metrics applied. There was large variation in network stability of contemporaneous structures across people, which could not be explained by subject-specific variables. We illustrate the temporal changes in contemporaneous network structures of two participants with high and low network stability and discuss the most pressing questions to be considered by future research.

*Keywords:* stability, replicability, consistency, idiographic, subject-specific, psychological networks, network comparison

*Word count:* 6803

Temporal stability of idiographic psychological networks

## Introduction

Mental disease remains a growing concern for public health, but there has been little progress in developing effective prevention and treatment strategies in the past decades. One possible explanation lies in the heterogeneity of symptom presentations of common mental disorders (Fried, 2017), which is not well addressed by traditional methods. One newly developed method to fill this need are idiographic network models. Subject-specific network models are a response to two recently voiced calls in clinical psychology. First, there seems to be a need for psychological research to re-orient towards idiographic methods that focus on intra-individual (idiographic) processes as opposed to group-level differences (Molenaar, 2004). Second, scholars have been proposing a paradigm shift away from reductionism towards studying the complexity of psychological phenomena (Borsboom, 2017; Fried & Robinaugh, 2020; Hofmann et al., 2016; Hofmann & Curtiss, 2018).

### Network theories and models in clinical psychology

The Network theory of mental disorders (Borsboom, 2017; Bringmann et al., 2013) attempts to integrate psychology with insights and methods from complexity science, offering a novel framework for understanding the underpinnings of psychopathology. It conceptualizes psychopathology as an emergent state of dynamically interacting elements, for example, psychological states, behaviors, or stressors. These elements are conceptualized as agents, meaning that they are mutually related in causal ways. Symptoms are thus thought to contribute, not result from, psychopathology. This account seems closely aligned with established clinical practices such as functional analysis in cognitive behavioral therapy, where a patient’s psychological disorder is visualized as a path diagram. These informal case conceptualizations describe the proposed mechanisms of a given disorder. They often feature dynamics which align with complex system behaviors, such as vicious cycles of dysfunctional thoughts and behaviors (Burger et al., 2020, 2021; Scholten et al., 2020). The recent attempts to formalize and study such symptom dynamics using network science thus fell on fertile ground.

Psychological network models (Epskamp, Borsboom, et al., 2018; Epskamp, Waldorp, et al., 2018) are the methodological workhorse that quantify and visualize such system structures and dynamics as nodes connected through pairwise edges. Nodes typically represent psychological variables, and edges represent their pairwise relationships. Relationships may be directed (granger-causal; see Molenaar, 2019) or undirected (correlational), positive or negative, and can differ in strength. Network models thus come in different flavors, pertaining to the estimation procedures by which edge parameters are modeled and derived, and lend themselves to a multitude of research questions. Besides traditional nomothetic (group-level) approaches, they are suited to study intra-individual processes using single-subject time-series data. Researchers in clinical psychology hope that such models could resolve what is known as the Therapist's dilemma, by which therapists need to make predictions about their individual patients based on research findings on the nomothetic that may not allow such inferences (Bastiaansen et al., 2019; Frumkin et al., 2020; Hoffart & Johnson, 2020; Jordan et al., 2020; Piccirillo & Rodebaugh, 2019; Rodebaugh et al., 2020).

Summing up, psychological network models have been described as "window into a patient's daily life" (Epskamp, van Borkulo, et al., 2018). The question remains whether this window provides an unobstructed and representative view. Is it merely a doorhole, or panoramic window? Are we getting a clear view inside, or are we mostly seeing our own reflections?

**Temporal stability of idiographic networks**

Concepts of stability and change lie at the heart of much clinical research. Why does a person change from healthy to depressed? Can therapy facilitate positive change beyond the duration of treatment? And if change occurs, what mechanism does it actually build on? Psychological networks offer new possibilities of mapping these questions onto idiographic research designs. For example, changes in network structure have been related to therapeutic progress (Thonon, Van Aubel, Lafit, Della Libera, & Larøi, 2020) or relapse (Wichers, Groot, & Psychosystems, 2016). Even subtle changes in network structure are of interest, as they are thought to act as potential early warning signals which may predict future major change, i.e. a system's phase transition from a healthy attractor state to a disordered one. For example, signs of critical slowing down, showing as increased auto-correlations and variances of items, may foreshadow a patient's relapse into depression upon stopping antidepressant treatment (Wichers et al., 2016); for similar work on resilience, see Kuranova et al., (2020). Such concepts are already successfully applied in other disciplines, for example ecology and climatology (Rusoja et al., 2018).

Single-subject research is quasi-experimental by nature (e.g., Beck & Jackson, 2021). The counterfactual question of "What would have happened otherwise?" cannot be answered as readily as in cross-sectional randomized controlled designs (e.g., Berlim et al., 2020). The individual remains their own control group, and it is essential to understand how this *control group* may or may not change when no *change* is expected. To our knowledge, only one study has investigated the temporal stability of idiographic psychological networks in a setting where no profound change was expected. Beck & Jackson (2020) investigated the consistency of idiographic personality over the course of two years. They found that personality networks of contemporaneous (at the same point of assessment), but not temporal (at assessments over time), associations were generally stable within individuals. They also found considerable interpersonal variability herein. In a quasi-experimental follow-up study investigating personality consistency in the context of the COVID-19 pandemic (Beck & Jackson, 2021), they arrived at similar conclusions and investigated potential antecedents that would explain variation of longitudinal network instability between subjects. Interpersonal variation appeared weakly linked to participants previous life satisfaction, but remained largely unexplained, from which the authors suggested that network instability might be considered a trait in itself. An alternative possibility, however, is that psychological network structures tend to naturally vary across time, meaning that instability may have occurred mainly due to within-person sampling variation.

**Aim of this study**

The present study aims to assess the stability of idiographic networks in the context of psychopathology from three different angles. First, we explore the temporal stability of networks within people. Second, we explore variation in network stability across commonly used stability metrics. Third, we explore potential predictors of temporal stability across people.

*RQ1*: How stable are estimated idiographic network structures over time?

*RQ2*: Do stability estimates vary across common stability metrics used?

*RQ3*: What person-specific or model-specific factors explain interpersonal variation in idiographic network stability?

We re-analyze daily diary data of people diagnosed with a personality disorder (Wright, Beltz, et al., 2015; Wright, Hopwood, et al., 2015). In summary, participants (N=112) provided once-daily ratings of their mood, behavior, and daily stressors over the course of one hundred consecutive days. We fit idiographic network models on participants' first (T1) and last (T2) fifty days of measurements to assess intra-individual stability (RQ1). We compare resulting network structures in their global structure on multiple stability metrics, and how these relate to each other (RQ2). Finally, we examine whether person-specific and network-specific factors explain interindividual variation in network stability using multivariate linear regression (RQ3). We illustrate and interpret the temporal network stability of two participants by example and end with a discussion of relevant questions to be addressed by future research.

## Methods

We used R (v4.0.5) for all our analyses. Raw data, estimated network objects, stability estimates and analysis scripts are available at github.com/RicardaP/thesis_repo and https://osf.io/gnw4s/.

### Data Set

Daily diary data was provided from people with a personality disorder (N=112). Over the course of 100 days, participants were instructed to keep nightly ratings of their mood, behavior, experiences of stress, and symptoms related to psychopathology. The study procedure and participant characteristics are described in detail by the original Authors (Wright, Hopwood, et al., 2015; Wright & Simms, 2016). We considered this data suited for our research questions for three main reasons. First, it is one of only a few data sets including comparably long time-series data in a relatively large sample. Second, the daily measurement intervals ensured that data would meet the assumption of equal distances between measurements, in contrast to more intensive time series data. Third, previous publications concluded that psychological processes in this sample were stable over the course of measurement (Wright & Simms, 2016), suggesting that network comparisons would less likely be confounded by unreliable measurements or major external events. Daily variables included in the current study are shown in Table 1.

**Table 1**

***Daily measures included in network models***

| Variable | Scale information |
|---|---|
| Positive Affect | "Please indicate to what extent you have felt this way over the past 24 hours: … “<br>Active, Alert, Attentive, Determined, Inspired<br>[Likert, 0 = very slightly to 4 = extremely] |

| | |
|---|---|
| Negative Affect | "Please indicate to what extent you have felt this way over the past 24 hours: … "<br>Afraid, Nervous, Hostile, Ashamed, Upset<br>[Likert, 0 = very slightly to 4 = extremely] |
| Stress | Seven items about stressful experiences, e.g., "Since this time yesterday, did anything happen that you could have argued about but you decided to let pass in order to avoid a disagreement? If yes, please rate the severity."<br>[Likert, 0 = not at all to 3 = very] |
| Task impairment | Single item: "How much difficulty did you have in taking care of important tasks or responsibilities?"<br>[Likert, 0 = not at all to 7 = extremely] |
| Behavior (other) | Single items: "Please judge how accurate each of the following words described your behavior over the past 24 hours:…" Dominant, Assertive, Critical, Irritable, Indifferent, Introverted, Passive, Submissive, Trusting, Warm, Sympathetic<br>[0 = extremely inaccurate : 5 = extremely accurate] |
| Personality disorder symptoms | Single items from the Daily Expression of Personality Disorders inventory (DPDS-32)<br>[0 = not at all : 7 = very much] |

*Note.* Positive Affect, Negative Affect, and Stress were included as composite scores and in each individual network, along with a single item Task impairment. Per network, two individual single item measures were included from the pool of behavior variables and personality disorder symptoms. Detailed item information is available under https://osf.io/ceaj2/.

## Network estimation

Contemporaneous and temporal idiographic networks of participants first (T1) and last (T2) fifty days of measurements were estimated as Gaussian Graphical Models (GGMs) using the R package *graphicalVAR* (v0.2.4, Epskamp, Waldorp, Mõttus, & Borsboom, 2018). Graphical VAR models belong to a wider family of partial correlation networks. Edges are estimated using vector autoregression (VAR). Graphical VAR estimates two types of networks: First, a temporal network of lagged effects is derived. Each variable in the network is modeled as a function of all other variables in the network at the previous lag (in our case, the previous day), including itself.

Edges are directed, can be positive or negative. Second, a contemporaneous network is derived from the residuals of the VAR model by inverting the variance-covariance matrix. This GGM captures contemporaneous, undirected associations between variables at the same measurement interval, controlling for any temporal associations estimated in through VAR. To prevent overfitting and aid interpretability, models are regularized using graphical least absolute shrinkage and selection operator (graphical LASSO, Friedman et al., 2008). Graphical LASSO regularizes parameter estimates according to a tuning parameter $\lambda$ that controls the sparsity of the network. The higher $\lambda$ , the more edge estimates will be shrunken to zero in order to optimize model fit and generalizability. Model fit is optimized according to the Bayesian Information Criterion (BIC) or extended BIC (EBIC, Chen & Chen, 2008), depending on how the hyperparameter $\gamma$ is specified. We estimated models using BIC ($\gamma = 0$) and $\lambda_{Min} = .025$, as done by Beck & Jackson (2020) and suggested by previous estimation studies (Epskamp, 2017). As a result, model estimation is less conservative in recovering smaller edges weights, which we considered appropriate for this exploratory research.

**Pre-processing**

Data were pre-processed in order to meet assumptions of the graphical VAR model. Along the usual assumptions pertaining to regression models, it assumes that data are measured at equal distances and without measurement error. Missing data points at thus pose a challenge for model estimation, as do highly skewed variables with little variance. We excluded participants whose responses were missing on more than 30 days in total, or more than 15 days at either T1 or T2. Remaining missing data were imputed by the Kalman Filter algorithm (Harvey, 1989) as implemented in the R package *imputeTS* (v3.2). Kalman imputation has been shown to recover idiographic network structures at levels up to 50% data missing completely at random (Mansueto et al., 2020). Graphical VAR modeling further assumes equal means and variances across time, known as the stationarity assumption. While many researchers note that this assumption may not be realistic in psychological data, it is generally recommended to transform data to by detrending effects of time 7/24/2021 2:17:00 AM. Imputation and detrending were performed at the level of the individual across the full 100 days. Variables were detrended independently of the magnitude and statistical sagnificance of the time effect.

**Variable selection for idiographic networks**

The diary data used in this study consists of a broad set of variables assessed in a comparably small and heterogeneous sample. Variables, shown in Table 1, were selected based on partly theoretical, partly statistical grounds according to three criteria: First, we strived for a set of variables that applied equally to all participants in the sample to make networks somewhat comparable across participants. Second, variance with more normal distributions were preferred to ensure model convergence. Third, we wanted to make the variable selection process reproducible and automizable. We included six predictors per network, as recommended by simulation work suggesting that graphical VAR performs well at such size for comparably small sample sizes (50 assessments per network) (Mansueto et al., 2020). Each network included three composite variables (mean scores) which were expected to fluctuate similarly across participants: positive affect (PA), negative affect (NA), and daily stress. We further included a single-item variable capturing daily functioning. We selected two additional items per subject according to the following ranking metric:

$$Ranking\ metric = (1 - SW)_{T1} * (1 - SW)_{T2} * proportion\ completed_{T1} * proportion\ completed_{T2}$$

The Shapiro-Wilk test statistic ($SW$) tests the null hypothesis that a variable is sampled from a normal distribution and ranging from 0 (close to normality) to 1, (Shapiro & Wilk, 1965). Capturing the items mean and variance in this way, we prioritized variables with minimal skew and maximal variance for the individual. These criteria were balanced against levels of missingness, because more missing and therefore imputed data would likely confound our network comparisons.

**Stability metrics**

We used three different metrics to assess temporal network stability across T1 and T2 (RQ1). Stability metrics were computed for contemporaneous and temporal networks.

**Edge weight correlations.** Global network similarity can be described as the correlation of estimated edge weights (Borsboom et al., 2017; Fried et al., 2018). Correlations can range from -1 to 1, with 0 indicating no relation between networks. This metric classifies structural similarity by the *strength and direction* of all pairwise relationships in their respective *order*, but is agnostic to the fact that edges may be zero due to regularization. When applied to regularized networks, as is the case here, it may therefore overestimate the similarity of sparse (i.e. largely unconnected or empty) structures.

**Jaccard similarity.** Jaccard similarity (as used in Borsboom et al., 2017) describes the proportion of replicated estimated edges (edges that are non-zero in both networks) out of all estimated but not necessarily replicated edges (edge that are non-zero in either network). As such, it classifies similarity by *whether* there is a pairwise relationship and is agnostic to the strength or direction of that relationship.

**Proportion of recovered edge signs.** To estimate the degree to which edge signs (i.e. positive or negative edge weights) replicate across networks, we calculated the proportion estimated edges with replicated sign out of all possible edges. This measure thus classifies similarity by the *sign* of the pairwise relationship, but is agnostic to the strength of the relationship, and may underestimate the similarity of sparse structures.

**Explaining inter-individual variance**

For contemporaneous network models, we explored what factors may explain intra-individual variance in network stability using multivariate linear regression (RQ2). Linear models were fit for each stability metric separately. Per metric, two sets of predictors were fit, leading to six linear models estimated. One set of predictors consisted of subject-specific factors measured at baseline, the other set captured statistical features of the diary data and network models. Predictor variables are listed in Table 2. These analyses were considered highly exploratory. Our choice of predictor variables was not based on prior expectations, as little research on this topic exists. The variables Happiness and life satisfaction were included to enable comparisons with Beck & Jackson (2021).

**Table 2**

***Predictor variables included in multivariate regression models***

| Variable | Description |
| --- | --- |
| *Set 1: Baseline variables related to subject-specific factors* | |
| Age | "What is your current age?"<br>[0 = female, 1 = male] |
| Gender | "What is your sex?" |
| Recent Treatment | "How long ago did the most recent treatment end?"<br>[0 = currently in treatment : 4 = 5 years ago, 5 = never been in treatment] |
| Happiness | "In the past 6 months, to what extent would you say you have been happy/satisfied/optimistic about yourself and your life?"<br>[0= not at all : 7=very much] |
| Life Satisfaction | Satisfaction With Life Scale (SWLS) |

| | |
|---|---|
| Big- 5 Personality traits | NEO - Five Factor Inventory (NEO-FFI) |
| *Set 2: Time-series variables related to network-specific factors* | |
| No. imputed | total number imputed item responses (sum T1 and T1) |
| Model fit | Bayesian Information Criterion (BIC) |
| Sparsity | proportion of empty edges (average T1 and T2) |
| Changes in Positive Affect | $\Delta_{T1-T2}(1 - SW_{PA})$ |
| Changes in Negative Affect | $\Delta_{T1-T2}(1 - SW_{NA})$ |
| Changes in Stress | $\Delta_{T1-T2}(1 - SW_{Stress})$ |
| Changes in Impairment | $\Delta_{T1-T2}(1 - SW_{Tasks})$ |

*Note*. SW: Shapiro-Wilk test statistic. T1: First 50 days of measurement. T2: Last fifty days of measurement.

## Results

### RQ1: Network stability within people

Of the original sample ($\mathrm{N_{original}} = 116$), four participants had been excluded by the original authors. Of the available 112 participants, twenty-seven had missed entries on more than 15 days in T1 or T2 each, and nine participants exhibited zero variance in daily task impairment. These participants were excluded from the analysis ($\mathrm{N_{excluded}} = 33$). Of the remaining participants ($\mathrm{N} = 79$), fifty-two participants were male. One person did not indicate their gender.

Networks included an average of 49 imputed data points (missing item responses) across T1 and T2 (Mdn = 43, SD =37.5), corresponding to an average of 8 missed days of assessment across the study period.

The most often included individual variables were DPDS19 (“I lied to someone”, 16 networks), DPDS8 (“I wanted people to notice my body”, 13 networks), DPDS31 (“I felt like I wanted to hurt someone”, 12 networks) and DPDS27 (“I behaved irresponsibly”, 12 networks).

Network models converged for all participants. Of the 79 estimated graphical VAR networks, 46 temporal network structures were empty and one contemporaneous network was empty.

Stability estimates were calculated for contemporaneous and temporal networks. Some stability metrics could not be calculated in subjects with empty networks at T1 and / or T2, so the number of derived estimates varies per metric and type of network, see Table 3.

Descriptives of the derived network comparisons are shown in Table 3. Stability metrics appear to be very unstable for temporal networks, as indicated by Median values of 0 across indices. Contemporaneous networks exhibited higher and more varying levels of stability across indices, with Median correlations of $r_{Spearman} = .48$, Median Jaccard similarity of .33, and a median of 13% of edges with equal signs at T1 and T2.

Figure 1 shows the distribution of stability indices for each contemporaneous network. Eight contemporaneous networks yielded negative edge weight correlations, and positive correlations appear to be evenly distributed across the full range of positive values. Jaccard similarities of contemporaneous networks span the full range as well, but appear to gravitate towards values below .50. Percentages of recovered edge signs of contemporaneous networks were mostly located between .00 and .25.

**Table 3**

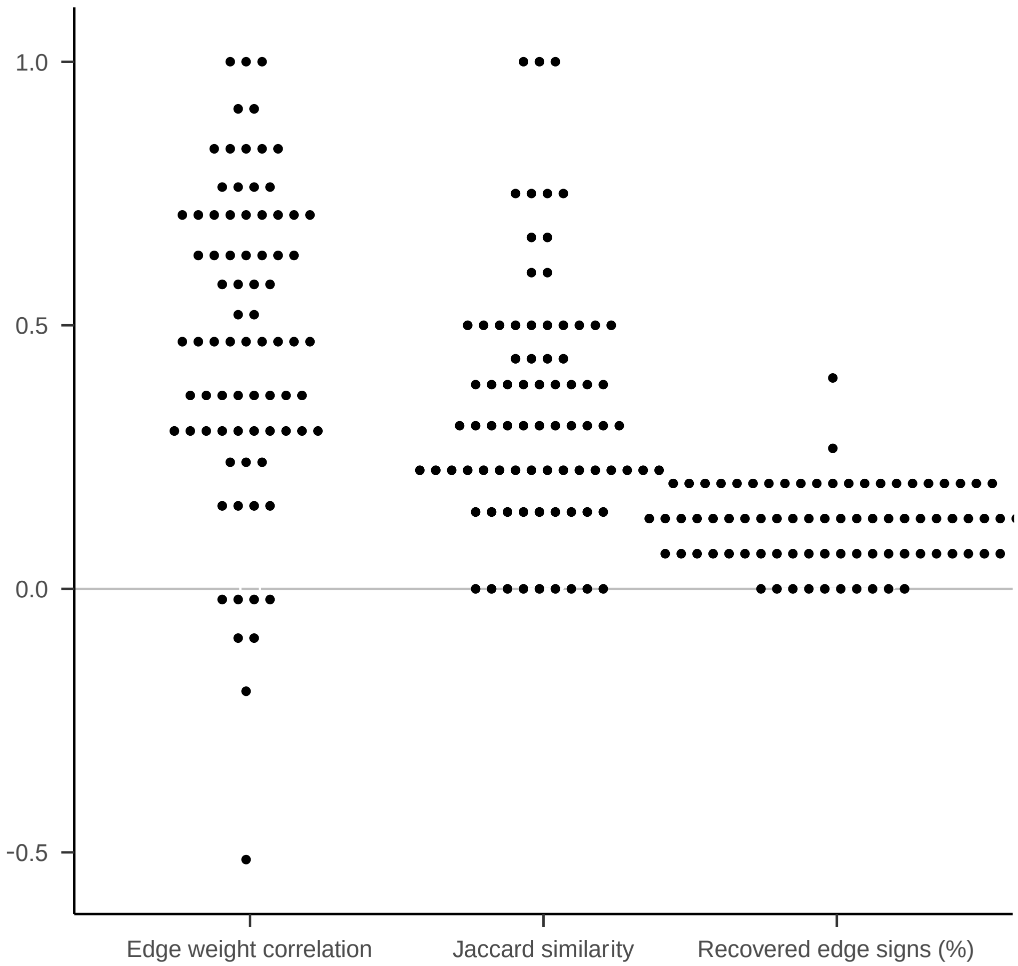
***Descriptive statistics of derived network stability estimates for temporal and contemporaneous networks***

| | Temporal networks | | | | Contemporaneous networks | | | |
|---|---|---|---|---|---|---|---|---|
| Metric | N | Mean | Median | Inter-quartile range | N | Mean | Median | Inter-quartile range |
| Edge weight correlation | 42 | 0.12 | 0.00 | -0.05 : 0.31 | 78 | 0.47 | 0.48 | 0.29 : 0.70 |
| Jaccard similarity | 70 | 0.08 | 0.00 | 0.00 : 0.18 | 78 | 0.34 | 0.33 | 0.20 : 0.50 |
| Replicated signs (%) | 79 | 0.01 | 0.00 | 0.00 : 0.03 | 79 | 0.12 | 0.13 | 0.07 : 0.20 |

Note. Number of derived estimates vary across metrics and type of network because some estimates cannot be calculated when networks are fully empty at T1 and / or T2.

**Figure 1**

Temporal stability contemporaneous idiographic networks




*Note.* Edge weight correlations can range from 1 to -1; Jaccard similarity and proportions of recovered edge signs from 0 to 1.

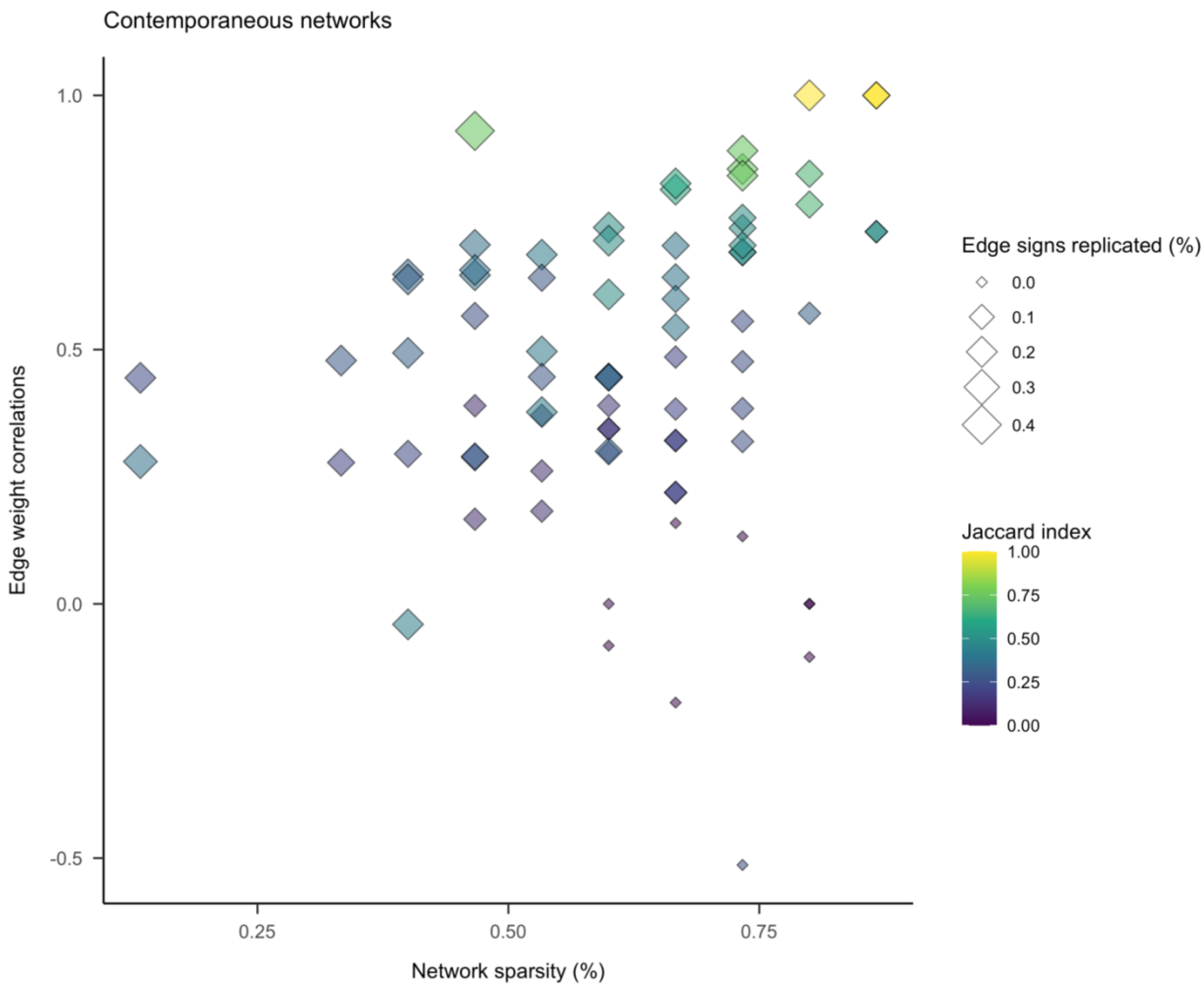
## RQ2: Network stability across metrics

As a high proportion of temporal networks were empty, and stability estimates of temporal networks were mostly zero, we focused on contemporaneous networks only for all further comparisons. Stability estimates within people displayed distinct distributions, which is not surprising given that they evaluate network similarity on different features. To understand how individual network structures are evaluated across metrics, we visualized metrics together as a function of network sparsity in Figure 2.

Figure 2 maps edge weight correlations as a function of network sparsity. Network sparsity described the proportion edge weights that were empty across T1or T2, meaning it is agnostic to whether edges are empty in either or both of the networks. Plotted elements represent each subject's location on the three stability metrics using location along the y-axis indicating edge weight correlation, color to indicate Jaccard similarity, and size indicating the proportion of edge signs replicated across T1 and T2. Higher temporal stability is thus indicated by higher position along the y-axis, more yellow shades of color, and bigger size of the element. Several things stood out by visual inspection, that can facilitate our understanding of how these metrics behave in relation to reach other. All metrics appeared to be related to network sparsity to varying degrees. Profile correlations appear positively related with network sparsity. They also appear to be spread more widely at higher levels of sparsity, as indicated by stronger positive *and* negative correlations at higher sparsity. A similar pattern could be observed for Jaccard similarity, where high values appeared predominantly at high levels of sparsity and high correlations. Proportions of edge signs replicated show more spread at high levels of sparsity as well, with the smallest values appearing in networks with negative edge weight correlations. This tendency is also visible in how the percentage of replicated edge signs behave as a function of sparsity, with the smallest estimates being located in sparse networks with negative edge weight correlations. A likely explanation for these observed patterns is that with fewer estimated edges in total, any change in sign or strength at one edge exerts proportionally more influence on the estimates of similarity of the global structure when the network is sparse.

**Figure 2**

*Stability metrics as a function of network sparsity*




*Note*. Network stability estimates as a function of network sparsity. Network sparsity describes the proportion edge weights that were empty across T1 or T2. Higher temporal stability is indicated by higher position along the y-axis (edge weight correlation), more yellow shades of color (Jaccard index), and bigger elements (proportion of edge signs replicated).

## RQ3: Network stability across people

To explore variables that may explain interpersonal variance in temporal stability of contemporaneous structures, we fit six linear regression models. First, we estimated models using each stability metric obtained for RQ2 as a dependent variable predicted by the set of baseline variables shown in Table 2. Next, each stability metric was predicted by the set of network-specific variables.

Of these six models, only the model predicting the proportion of replicated edge signs by network-specific variables showed significant ($\alpha < .05$) model fit, $R^2 = .22$, $F(7, 71) = 2.89$, $p < 0.01$. In this model, sparsity was the only significant predictor of network stability, $\beta = -.47$, $t = -4.33$, $p < .001$. This means that lower sparsity was negatively related to proportion of replicated edge signs. Note that the proportion of replicated edge signs is a metric directly affected by the number of empty edges in a network.

Two models showed noteworthy ($\alpha < .10$) model fit when predicted by baseline variables. With edge weight correlations as the dependent variable, Life satisfaction was positively related to network stability, $\beta = .36$, $t = 2.4$, $p = .019$. Conscientiousness was negatively related to edge weight correlations, $\beta = -.49$, $t = -3.16$, $p = .002$. Model fit was not significant $R^2 = .23$, $F(10,66) = 1.95$, $p = .054$. With Jaccard similarity as the dependent variable, Conscientiousness appeared to be negatively related to Jaccard similarity, $\beta = -.39$, $t =-2.54$, $p = .014$. Agreeableness appeared to be positively related to Jaccard similarity, $\beta = .25$, $t = 2.06$, $p = .044$. Model fit was not significant, $R^2 = .21$, $F(10, 67) = 1.79$, $p = .079$. Results should be interpreted with caution, but we highlight them here because these are variables future investigations may want to focus on.

Three remaining models explained less than 15% of variance in outcome (all $p > .10$) variables and significant predictors in these were disregarded. The model predicting proportion of edge signs replicated by baseline variables explained 9% of variance, $R^2= 0.09$, $F(10,67) = 0.72$, $p=.705$. The model predicting edge weight correlations by network-specific variables explained 6% of variance, $R^2= .06$, $F(7, 70)=.66$, $p = .705$. The model predicting Jaccard similarity by network specific variables explained 11 % of variance, $R^2 = .11$, $F(7, 71) = 1.28$, $p =.272$.

**Examples**

Two participants were selected for illustration. We selected these participants to exemplify cases of relatively high and low temporal stability, while ensuring other potentially confounding features were similar. Participant 53 (P35) and participant 45 (P45) satisfied these criteria. Their networks contained only 2 and 3 imputed data points, respectively. Their contemporaneous networks had similar levels of sparsity ($Sparsity_{P53} = .53$; $Sparsity_{P45} = .40$), and their model fit was most similar given the previous criteria ($Mean_{BIC\ P53} = 196.66$; $Mean_{BIC\ P45} = 236.66$).

The networks of participant 53 exhibited relatively low temporal stability. The similarity of global network structure as measured by edge weight correlation was low, $r_{Spearman\ P53} = .18$. Jaccard similarity was low, $Jaccard_{P53} = .14$. Only one of the six estimated edges had the same sign at T1 and T2. In contrast, stability metrics for participant 45 suggest high temporal stability, with edge correlation $r_{\text{Spearman P45}} = .648$ , $Jaccard_{P45} = .33$ , and *replicated edge signs* $(\%)_{P45} = .2$ (three edges with equal sign across T1 and T2).

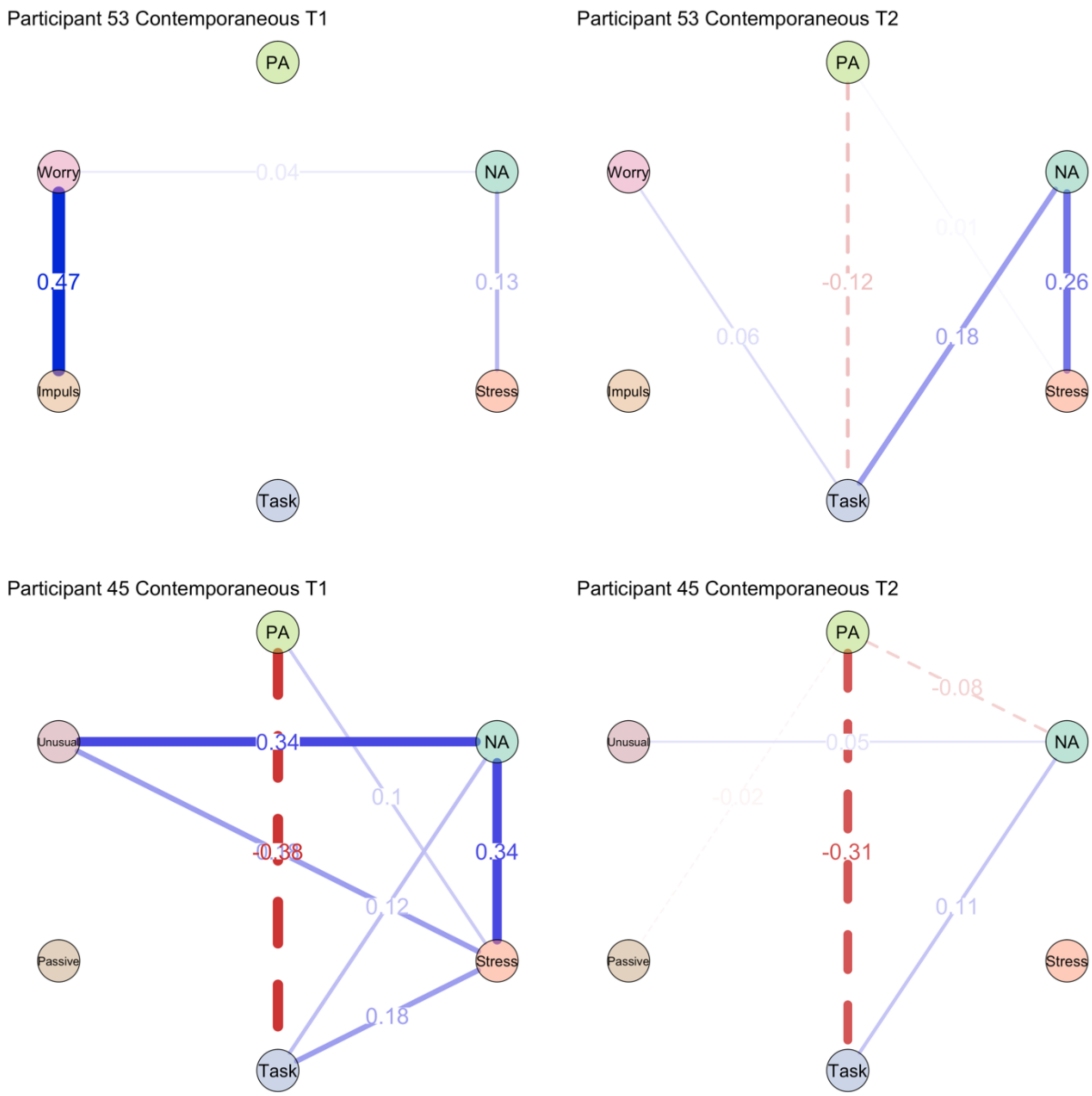
Figure 3 shows their estimated contemporaneous networks at T1 and T2. Their scaled daily item ratings across T1 and T2 are shown in Figure 4. Means, standard deviations, and changes in distributions of raw daily measures for participants 53 and 45 are shown in Table 4. Individual variables selected for each person's network according to the rank score explained before were DPD25 ("I acted on impulse while feeling upset") and DPD21 ("I worried about being abandoned") for participant 53, and "Passive behavior" and DPD24 ("I did things others may find unusual") for participant 45.

**Participant 53 with low stability of contemporaneous network.** For this participant, there was a strong positive contemporaneous association of Worry and Impulse at T1, which was not present at T2. The only association that was found at T1 and T2 was the positive association of Negative Affect and Stress, this association was stronger at T2 than at T1. At T1. Worry and Negative Affect were weakly positively related. Interestingly, this direct association was not present at T2, but an indirect association occurred via Task Impairment. Task impairment was positively related to Worry and Negative affect, and showed a negative association with positive affect. At T2, there was an additional weak positive relation between Positive affect and Stress. Item means and variances over time shown in Figure 4 show that overall, this person appeared to experience frequent fluctuations on all variables across T1 and T2, which appear slightly less frequent at T2.

**Participant 45 with high stability of contemporaneous network.** The contemporaneous networks of Participant 45 were generally denser than for Participant 53. At T1, there were strong positive associations between Negative affect and Stress, and Negative Affect and Unusual behavior. Negative affect as also related to task impairment, although somewhat weaker. At T2, the strong positive edge between negative affect and stress disappeared, the edges between negative affect and task impairment remained very similar, but the edge between negative affect and unusual behavior, was considerably smaller compared to T1. Positive affect showed a strong negative association with task impairment at both T1 and T2. At T1, positive affect was also weakly related with Stress. This connection was not present at T1, where positive affect showed two previously absent negative relations with negative affect and passive behavior. Interestingly, Stress was a highly connected variable in the network at T1, with positive associations to Negative affect, positive affect, unusual behavior and task impairment. At T2, stress was not related to any other variable in the network. Item means and variances over time shown in Figure 4 suggest that stress was much higher and more variable at T1 for this participant. They also seemed to experience fewer peaks of negative affect and passive or unusual behavior, while positive affect appeared to generally high and stable across T1 and T2, see Table 4.

**Figure 3**

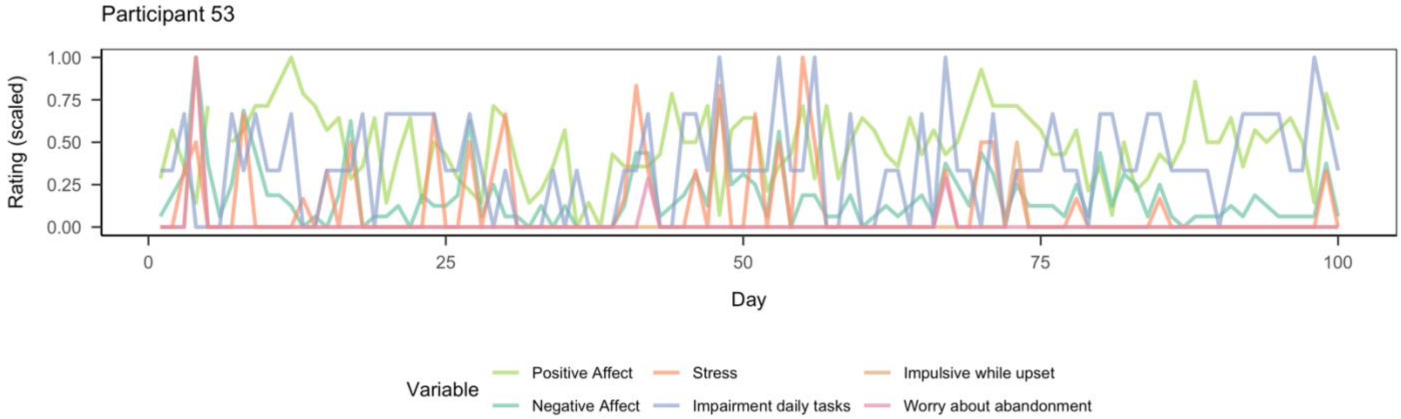
*Contemporaneous network structures of Participant 53 and 45 at T1 and T2.*

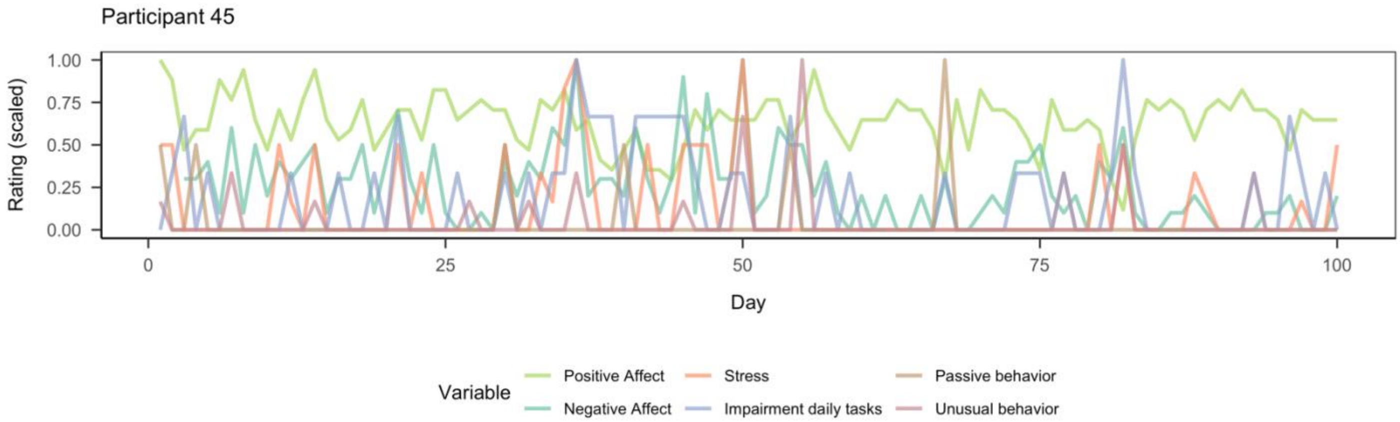



*Note*. Maximum edge width set to .47 across networks, so that they can be compared across subjects. Edges reflect contemporaneous partial correlations as estimated by graphical VAR using BIC ($\gamma = 0$) and $\lambda_{Min} = .025$. PA: Positive Affect. NA: Negative effect. Task: Impairment on daily tasks. Worry: Worrying about being abandoned. Impulse: Acted impulsive while upset. Passive: Passive behavior. Unusual: Doing things others may find unusual.

***Figure 4***

*Scaled daily rating of variables in cluded in each participants network*




*Note.* Daily Ratings of variables included in participant's networks. Ratings are scaled to reflect the individual's maximum range on each item separately. Positive Affect, Negative Affect, and Stress are composite variables, all other variables are single item ratings.

***Table 4***
**Means, standard deviations, and changes in distributions of raw daily measures.**

| Participant | Variable | $Mean(SD)_{T1}$ | $Mean(SD)_{T2}$ | $(1 - SW)_{\Delta T1-T2}$ |
|---|---|---|---|---|
| P53 | Positive Affect | 1.27 (0.67) | 1.41 (0.51) | 0.01 |
| | Negative Affect | 0.66 (0.71) | 0.49 (0.40) | 0.03 |
| | Stress | 0.12 (0.21) | 0.09 (0.19) | -0.10 |
| | Task Impairment | 1.00 (0.83) | 1.22 (0.86) | 0.03 |
| | Impulsive while upset (DPDS25) | 0.08 (0.56) | 0.04 (0.28) | 0.00 |
| | Worry about abandonment (DPDS21) | 0.18 (1.02) | 0.04 ( 0.28) | -0.05 |
| P45 | Positive Affect | 2.18 (0.55) | 2.16 (0.50) | -0.09 |
| | Negative Affect | 0.69 (0.49) | 0.32 (0.35) | -0.07 |
| | Stress | 0.16 (0.24) | 0.05 (0.12) | -0.25 |
| | Task Impairment | 0.72 (0.85) | 0.39 (0.67) | -0.15 |
| | Passive behavior | 0.08 (0.27) | 0.06 (0.31) | -0.11 |
| | Unusual behavior (DPDS24) | 0.26 (0.72) | 0.18 (0.95) | -0.22 |

$Note.$ $(1 - SW)_{\Delta T1-T2}$: Changes in distributions of raw variables, calculated as

(1-Shapiro Wilk test statistic $)_{T1}$ - (1-Shapiro Wilk test statistic$)_{T2}$ .

## Discussion

This study was the first to systematically explore the temporal stability of idiographic psychopathology networks, and differences herein, in a setting where temporal stability could be expected. Stability was assessed using three metrics of network similarity, and their findings were compared. Estimated temporal networks structures appeared to be highly unstable over time. Estimated contemporaneous networks appeared to be moderately stable within subjects over time. For contemporaneous networks, the assessment of stability varied across metrics applied. There was also large variation in network stability across people, which could not be explained well by subject-specific or network-specific variables as regression model fit was poor. Life satisfaction and agreeableness appeared to be positively related to network stability, and conscientiousness appeared to be negatively related to network stability. However, overall model fit was not significant, and these results should therefore be interpreted with caution. Further, these patterns were not consistent across outcome measures and should be regarded purely exploratory findings.

Our results echo findings of two previous studies on the temporal stability of idiographic personality networks (Beck & Jackson, 2020, 2021). Similar to both of these studies, temporal network structures were dramatically less stable compared to contemporaneous network structures. Beck & Jackson speculate that the lack of stability in temporal associations may be owed to a mismatch of the modeled time interval, and the timing at which variables can be expected to affect each other (Epskamp, van Borkulo, et al., 2018). In our study, temporal associations were estimated one day apart, and it seems plausible that any temporal association may be due to more situational aspects causing covariance across lags, such as, acute stressors lasting multiple days. Stability estimates for contemporaneous networks found in our study were on average lower than those reported by Beck & Jackson (2020), but span a similar range of values including negative correlations, where some people's networks remained highly stable, some people's structures were very unstable, and a few people's structures showed opposite patterns in form of systematically different signs in estimated edges.

The source of interpersonal differences in network stability is unclear. Our exploratory analyses suggested three candidate variables to be investigated in future research: life satisfaction, agreeableness, and conscientiousness. Life satisfaction showed a weak positive relation with edge weight correlations, bearing in mind non-significant model fit. This finding is in contrast with Beck & Jackson, 2021, who reported weak positive associations of satisfaction with network stability during COVID-19. Beck & Jackson interpret their findings in terms of adaption and adjustment, offering two competing explanations. High life satisfaction may be a sign of people's ability to adapt to new circumstances and adjust their behavioral patterns in response to a pandemic. Alternatively, people who were satisfied prior to the pandemic may have experienced the situational changes as more disruptive. Our preliminary finding of a positive association of life satisfaction at consecutive times put these speculations in a different light. Possibly, in absence of major external disruptions, well-adapted people with high life-satisfaction fare well in sticking to their adaptive patterns (high network stability), whereas people with less life satisfaction may continuously experiment with their ways of behaving in order to arrive at an adaptive set of behaviors. While offering an intuitive argumentation for why life satisfaction may be positively and negatively related to network stability, the argument hinges on conceptualizing people with low life satisfaction as maladapted and rigid in a pandemic scenario, and as maladapted and erratic in a business-as-usual scenario, for which no intuitive explanation comes to mind.

Despite these open questions, we believe that one conclusion can be drawn to summarize the above findings: The question is not whether psychological networks are stable within people, but when, for who and to what degree. However, we see some major roadblocks ahead of this scientific quest. Therefore, we outline the most pressing questions to be tackled next.

**Sampling variability**

One important assumption made in this work, and by research using graphical VAR, is that data is measured without error. Given that psychological measurement has plagued researchers for decades, this may be a strong assumption to make (Hallquist et al., 2021; Schuurman & Hamaker, 2019). We tried to mitigate this in our study by using three composite alongside three single-item measures. However, measurement error is not the only source of noise that may affect the conclusions drawn from network analysis. Investigations on idiographic network stability are particularly challenged by noise in parameter estimates, as there are currently no readily available methods to assess whether a given edge weight is an unbiased estimate of the true relationship in a larger sample (Epskamp, Borsboom, et al., 2018; Forbes et al., 2021a; Fried et al., 2021). Similar to coefficient estimates in linear regression, the coefficient estimates inform us about the strength of a relationship *in the data at hand*, but they tell us little about how dependent that estimate is on the particular sample. Methods to assess the stability of network parameters are readily implemented for networks using group-based estimation procedures, but they have yet to be developed for single-subject data.

**Stability assessment.**

Returning to the counterfactual question of "How stable are networks when stability can be assumed?", we need to ask whether a degree of stability indicated signals a meaningful difference, or merely sampling variability from the ups and downs of daily life. There have been debates about what network features should be considered relevant in comparison of networks (Borsboom et al., 2017; Forbes et al., 2021b). In the current study, we saw that different metrics favor different features of a network, and they may disagree at their extreme ends. For example, the sparser the network, the easier it was classified as highly stable or highly unstable by the similarity metrics edge correlations, Jaccard index, and proportion of replicated edge signs. This is not surprising, as simplification is somewhat the point of applying regularization. However, an empty edge in a regularized network is no evidence of conditional independence (Epskamp et al., 2017), and commonly used stability metrics such as edge correlations and Jaccard similarity are agnostic to regularization. Furthermore, some metrics are sensitive to changes in estimated edge signs (e.g., edge weight correlations) whereas others, like Jaccard index, are not. We believe that the simulation and reanalysis work currently done on this topic would strongly benefit from empirical work to provide meaningful anchor points that map similarity metrics to actual outcomes. Ultimately, the question whether an edge weight correlation of .5 should be considered stable or instable can barely be answered by numbers alone, as the answer depends on how these values behave in relation to substantive outcomes.

**Stationarity**

The effects data imputation and linear detrending have on estimated network structures are not well known, and some authors report vastly different resulting network structures depending on what transformation was applied (Bastiaansen et al., 2019; Vos et al., 2017). Detrending a variable changes variance and distribution, because detrended scores reflect deviations from a variable's trend over time (the residuals of linear regression model predicting variable scores by time). Because graphical VAR modeling performs variance-covariance decomposition, changes in variance may result in changes in power and potentially biased path estimates if the assumption of stationarity does not hold. Part of our motivation for this research was studying whether the assumption of stationarity is or is not realistic for idiographic data. As examining differences in network structures both within and between people was the main goal of our study, we tried to equalize pre-processing decisions across variables and participants while working in an idiographic framework. However, it remains unclear whether linear detrending merely enables model estimation, or may also distort it (Bastiaansen et al., 2019).

One concern we see for applying linear detrending procedures is that trends in typical psychological data cannot be linear over extended periods of time. By definition, items are usually bounded within limits, and any trend estimated at a given window would exceed those limits when extrapolated. It seems more plausible that non-stationary trends follow a wave-like pattern over time, oscillating between item bounds at different amplitudes and phases. Note that the assumption of stationarity itself cannot be avoided for this type of model, but time-varying VAR models are being developed (Bringmann et al., 2018; Haslbeck et al., 2021). In these models, non-stationarity is estimated as part of the network itself, and linear as well as non-linear trends can be taken into account.

**Conclusion**

Are psychological networks a window into the patient's life? Yes. Is it a doorhole or a panoramic window? It depends. But for the near future, we may be well advised to further sharpen our vision before trusting our eyes blindly.

**References**

Bastiaansen, J. A., Kunkels, Y. K., Blaauw, F., Boker, S. M., Ceulemans, E., Chen, M., Chow, S.-M., Jonge, P. de, Emerencia, A. C., Epskamp, S., Fisher, A. J., Hamaker, E. L., Kuppens, P., Lutz, W., Meyer, M. J., Jr, R. G. M., Oravecz, Z., Riese, H., Rubel, J. A., … Bringmann, L. (2019). *Time to get personal? The impact of researchers' choices on the selection of treatment targets using the experience sampling methodology*. https://doi.org/10.31234/osf.io/c8vp7

Beck, E. D., & Jackson, J. J. (2020). Consistency and change in idiographic personality: A longitudinal ESM network study. *Journal of Personality and Social Psychology*, *118*(5), 1080–1100. https://doi.org/10.1037/pspp0000249

Beck, E. D., & Jackson, J. J. (2021). Idiographic personality coherence: A quasi experimental longitudinal ESM study. *European Journal of Personality*, 08902070211017746. https://doi.org/10.1177/08902070211017746

Berlim, M. T., Richard-Devantoy, S., dos Santos, N. R., & Turecki, G. (2020). The network structure of core depressive symptom-domains in major depressive disorder following antidepressant treatment: A randomized clinical trial. *Psychological Medicine*, 1–15. https://doi.org/10.1017/S0033291720001002

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375

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

Bringmann, L. F., Ferrer, E., Hamaker, E. L., Borsboom, D., & Tuerlinckx, F. (2018). Modeling Nonstationary Emotion Dynamics in Dyads using a Time-Varying Vector-Autoregressive Model. *Multivariate Behavioral Research*, *53*(3), 293–314. https://doi.org/10.1080/00273171.2018.1439722

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A Network Approach to Psychopathology: New Insights into Clinical Longitudinal Data. *PLOS ONE*, *8*(4), e60188. https://doi.org/10.1371/journal.pone.0060188

Burger, J., Epskamp, S., Veen, D. C. van der, Dablander, F., Schoevers, R. A., Fried, E. I., & Riese, H. (2021). *A clinical PREMISE for personalized models: Towards a formal integration of case formulations and statistical networks*. PsyArXiv. https://doi.org/10.31234/osf.io/bdrs7

Burger, J., van der Veen, D. C., Robinaugh, D. J., Quax, R., Riese, H., Schoevers, R. A., & Epskamp, S. (2020). Bridging the gap between complexity science and clinical practice by formalizing idiographic theories: A computational model of functional analysis. *BMC Medicine*, *18*(1), 99. https://doi.org/10.1186/s12916-020-01558-1

Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*(3), 759–771. https://doi.org/10.1093/biomet/asn034

Epskamp, S. (2017). Brief Report on Estimating Regularized Gaussian Networks from Continuous and Ordinal Data. *ArXiv:1606.05771 [Stat]*. http://arxiv.org/abs/1606.05771

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

Epskamp, S., Kruis, J., & Marsman, M. (2017). Estimating psychopathological networks: Be careful what you wish for. *PLOS ONE*, *12*(6), e0179891. https://doi.org/10.1371/journal.pone.0179891

Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A.-M., Riese, H., & Cramer, A. O. J. (2018). Personalized Network Modeling in Psychopathology: The Importance of Contemporaneous and Temporal Connections. *Clinical Psychological Science*, *6*(3), 416–427. https://doi.org/10.1177/2167702617744325

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). 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

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2021a). On Unreplicable Inferences in Psychopathology Symptom Networks and the Importance of Unreliable Parameter Estimates. *Multivariate Behavioral Research*, *56*(2), 368–376. https://doi.org/10.1080/00273171.2021.1886897

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2021b). Quantifying the Reliability and Replicability of Psychopathology Network Characteristics. *Multivariate Behavioral Research*, *56*(2), 224–242. https://doi.org/10.1080/00273171.2019.1616526

Fried, E. I. (2017). Moving forward: How depression heterogeneity hinders progress in treatment and research. *Expert Review of Neurotherapeutics*, *17*(5), 423–425. https://doi.org/10.1080/14737175.2017.1307737

Fried, E. I., Borkulo, C. D. van, & Epskamp, S. (2021). On the Importance of Estimating Parameter Uncertainty in Network Psychometrics: A Response to Forbes et al. (2019). *Multivariate Behavioral Research*, *56*(2), 243–248. https://doi.org/10.1080/00273171.2020.1746903

Fried, E. I., Eidhof, M. B., Palic, S., Costantini, G., Huisman-van Dijk, H. M., Bockting, C. L. H., Engelhard, I., Armour, C., Nielsen, A. B. S., & Karstoft, K.-I. (2018). Replicability and Generalizability of Posttraumatic Stress Disorder (PTSD) Networks: A Cross-Cultural Multisite Study of PTSD Symptoms in Four Trauma Patient Samples. *Clinical Psychological Science*, *6*(3), 335–351. https://doi.org/10.1177/2167702617745092

Fried, E. I., & Robinaugh, D. J. (2020). Systems all the way down: Embracing complexity in mental health research. *BMC Medicine*, *18*(1), 1–4. https://doi.org/10.1186/s12916-020-01668-w

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

Frumkin, M. R., Piccirillo, M. L., Beck, E. D., Grossman, J. T., & Rodebaugh, T. L. (2020). Feasibility and utility of idiographic models in the clinic: A pilot study. *Psychotherapy Research*, *0*(0), 1–15. https://doi.org/10.1080/10503307.2020.1805133

Hallquist, M. N., Wright, A. G. C., & Molenaar, P. C. M. (2021). Problems with Centrality Measures in Psychopathology Symptom Networks: Why Network Psychometrics Cannot Escape Psychometric Theory. *Multivariate Behavioral Research*, *56*(2), 199–223. https://doi.org/10.1080/00273171.2019.1640103

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Haslbeck, J. M. B., Bringmann, L. F., & Waldorp, L. J. (2021). A tutorial on estimating time-varying vector autoregressive models. *Multivariate Behavioral Research*, *56*(1), 120–149. https://doi.org/10.1080/00273171.2020.1743630

Hoffart, A., & Johnson, S. U. (2020). Within-person networks of clinical features of social anxiety disorder during cognitive and interpersonal therapy. *Journal of Anxiety Disorders*, *76*, 102312. https://doi.org/10.1016/j.janxdis.2020.102312

Hofmann, S. G., & Curtiss, J. (2018). A complex network approach to clinical science. *European Journal of Clinical Investigation*, *48*(8), e12986. https://doi.org/10.1111/eci.12986

Hofmann, S. G., Curtiss, J., & McNally, R. J. (2016). A Complex Network Perspective on Clinical Science. *Perspectives on Psychological Science*, *11*(5), 597–605. https://doi.org/10.1177/1745691616639283

Jordan, D. G., Winer, E. S., & Salem, T. (2020). The current status of temporal network analysis for clinical science: Considerations as the paradigm shifts? *Journal of Clinical Psychology*, *76*(9), 1591–1612. https://doi.org/10.1002/jclp.22957

Kuranova, A., Booij, S. H., Menne-Lothmann, C., Decoster, J., van Winkel, R., Delespaul, P., De Hert, M., Derom, C., Thiery, E., Rutten, B. P. F., Jacobs, N., van Os, J., Wigman, J. T. W., & Wichers, M. (2020). Measuring resilience prospectively as the speed of affect recovery in daily life: A complex systems perspective on mental health. *BMC Medicine*, *18*(1), 36. https://doi.org/10.1186/s12916-020-1500-9

Mansueto, A. C., Wiers, R., van Weert, J., Schouten, B. C., & Epskamp, S. (2020). *Investigating the Feasibility of Idiographic Network Models*. https://doi.org/10.31234/osf.io/hgcz6

Molenaar, P. C. M. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement:*

*Interdisciplinary Research & Perspective*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Molenaar, P. C. M. (2019). Granger Causality Testing with Intensive Longitudinal Data. *Prevention Science*, *20*(3), 442–451. https://doi.org/10.1007/s11121-018-0919-0

Piccirillo, M. L., & Rodebaugh, T. L. (2019). Foundations of idiographic methods in psychology and applications for psychotherapy. *Clinical Psychology Review*, *71*, 90–100. https://doi.org/10.1016/j.cpr.2019.01.002

Rodebaugh, T. L., Frumkin, M. R., & Piccirillo, M. L. (2020). The long road from person-specific models to personalized mental health treatment. *BMC Medicine*, *18*(1), 365. https://doi.org/10.1186/s12916-020-01838-w

Rusoja, E., Haynie, D., Sievers, J., Mustafee, N., Nelson, F., Reynolds, M., Sarriot, E., Swanson, R. C., & Williams, B. (2018). Thinking about complexity in health: A systematic review of the key systems thinking and complexity ideas in health. *Journal of Evaluation in Clinical Practice*, *24*(3), 600–606. https://doi.org/10.1111/jep.12856

Scholten, S., Lischetzke, T., & Glombiewski, J. (2020). *Toward Data-Based Case Conceptualization: A Functional Analysis Approach With Ecological Momentary Assessment*. PsyArXiv. https://doi.org/10.31234/osf.io/prg7n

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods*, *24*(1), 70–91. https://doi.org/10.1037/met0000188

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52*(3/4), 591–611. https://doi.org/10.2307/2333709

Vos, S. de, Wardenaar, K. J., Bos, E. H., Wit, E. C., Bouwmans, M. E. J., & Jonge, P. de. (2017). An investigation of emotion dynamics in major depressive disorder patients and healthy persons using sparse longitudinal networks. *PLOS ONE*, *12*(6), e0178586. https://doi.org/10.1371/journal.pone.0178586

Wichers, M., Groot, P. C., & Psychosystems, ESM Group, EWS Group. (2016). Critical Slowing Down as a Personalized Early Warning Signal for Depression. *Psychotherapy and Psychosomatics*, *85*(2), 114–116. https://doi.org/10.1159/000441458

Wright, A. G. C., Beltz, A. M., Gates, K. M., Molenaar, P. C. M., & Simms, L. J. (2015). Examining the Dynamic Structure of Daily Internalizing and Externalizing Behavior at Multiple Levels of Analysis. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01914

Wright, A. G. C., Hopwood, C. J., & Simms, L. J. (2015). Daily Interpersonal and Affective Dynamics in Personality Disorder. *Journal of Personality Disorders*, *29*(4), 503–525. https://doi.org/10.1521/pedi.2015.29.4.503

Wright, A. G. C., & Simms, L. J. (2016). Stability and fluctuation of personality disorder features in daily life. *Journal of Abnormal Psychology*, *125*(5), 641–656. https://doi.org/10.1037/abn0000169