



OPEN Dynamic time warp versus vector autoregression models for network analyses of psychological processes

Floor van der Does¹, Wessel van Eeden¹, Laura F. Bringmann², Femke Lamers³, Brenda W. J. H. Penninx³, Harriëtte Riese⁴, Eric Vermetten¹, Nic van der Wee¹ & Erik Giltay^{1,5}✉

Psychopathological disorders are increasingly conceptualized as complex dynamic systems, which can be represented as networks of interconnected symptoms. These dynamic networks are often constructed using Multilevel Vector AutoRegression (mIVAR) models. However, psychological processes frequently violate these assumptions. An alternative approach for examining temporal relationships between variables is Dynamic Time Warping (DTW). This paper evaluates the potential applications, advantages, and disadvantages of DTW and mIVAR. As part of the Netherlands Study of Depression and Anxiety, an Ecological Momentary Assessment module was administered five times daily for 2 weeks, to 376 participants (Mean age 49.3 years, 64.4% women). We created item networks based on 20 of the mood and physical condition items from this module using the mIVAR and DTW techniques, and repeated these analyses using simulated data to explore violations of mIVAR assumptions, including various lagged relationships and the presence of collider variables. Analysis of simulated datasets revealed that mIVAR networks were more susceptible to spurious connections, while DTW produces more reliable networks under these conditions. While mIVAR better reveals causal relationships when assumptions are met, DTW provides a robust method for examining co-occurrence, synchrony, and the directionality of lagged connections in real-world psychological data.

In most areas of medicine, symptoms of pathology typically originate from an identifiable common cause, an underlying disorder. For instance, in the case of COVID-19, most symptoms arise directly from the underlying viral infection, which can be identified separately from its symptomatic expression. Traditionally, psychiatric disorders have been conceptualized in a similar manner, where symptoms such as irritability, insomnia, and changes in appetite are considered direct consequences of an underlying disorder, such as Major Depressive Disorder¹. However, this common-cause model encounters substantial challenges within psychopathology. Unlike somatic illnesses, many psychiatric disorders are solely defined by the presence of a specific combination of symptoms. Consequently, an asymptomatic depression is impossible, and depression cannot be identified separately from its symptoms². This creates a circular reasoning wherein symptoms of a disorder are both the cause and the consequence of the disorder itself. Moreover, progress in identifying specific biological and environmental causes for various psychopathological conditions has been limited^{1,3}. This suggests that the conventional model of psychiatric disorders may be insufficient¹. In research, and to a lesser extent in clinical contexts, there is a prevailing focus on the presence or absence of diagnoses of entire syndromes⁴, even though this provides relatively limited clinical information due to the vast heterogeneity in symptom presentations that could meet criteria for diagnosis. Furthermore, individuals who do not meet the full criteria for diagnosis may still experience substantial distress⁵. Also, clinical diagnoses or conclusions about symptom severity are often based on single, or relatively few, measurement points, failing to consider the complex and dynamic nature of mental health conditions⁶. A more nuanced and comprehensive approach is needed to capture the intricate interplay between symptoms, biology, and the environment in the development and persistence of psychiatric disorders.

¹Department of Psychiatry, Leiden University Medical Center (LUMC), Leiden, The Netherlands. ²Department of Psychometrics and Statistics, University of Groningen, Groningen, The Netherlands. ³Department of Psychiatry, Amsterdam University Medical Center, Vrije Universiteit, Amsterdam, The Netherlands. ⁴Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. ⁵Department of Public Health & Primary Care, Health Campus the Hague, Leiden University, The Hague, The Netherlands. ✉email: e.j.giltay@lumc.nl

In recent years, the field of psychiatry has undergone a notable shift in its conceptualization of mental disorders and embraced complex dynamic systems theory⁷. This alternative framework posits that psychiatric disorders are constellations of interconnected symptoms that mutually influence each other, rather than a collection of symptoms that merely co-occur due to a shared origin². Symptoms and their mutual influences can be modelled in networks, which can include dynamic feedback loops and vicious cycles. For instance, negative mood and loss of energy may contribute to concentration difficulties, increased worrying, and sleep disturbances, which, in turn, can exacerbate the loss of energy and intensify negative mood¹. As a result, the disorder can become self-sustaining and persist even when any initial identifiable cause, such as a stressful experience or challenging life circumstances, has subsided^{2,7}. This theory allows a focus on the structure and dynamics of psychological complaints, instead of focussing on classifying individuals as “cases” or “non-cases”². This could improve our understanding of psychopathology, and facilitate personalized care⁸.

The dynamic interplay between symptoms or mood states can be studied and mapped in a network, using time-series data. Ecological Momentary Assessment (EMA), wherein participants complete brief questionnaires multiple times a day using a mobile device, may be an especially valuable approach. EMA provides an opportunity to capture detailed real-time information on mood states and their interactions, offering insights into subtle nuances and connections that may not be captured when relying on less frequent and retrospective assessments⁹. When constructing EMA networks, researchers often employ the Multi-Level Vector Autoregression (mlVAR) model. This model enables the fitting of an autoregressive model across multiple individuals, where the value of a symptom or mood state is predicted by its own value and the value of other items in time-lag analyses¹⁰. When researching psychopathology, it is common to use a lag-1 model, which regresses the values of the item scores over values from the previous time point^{11,12}. Findings are explained in terms of Granger causality, meaning that at least the temporal sequence requirements of causal effects are met, and fluctuation in a ‘causal’ item precedes fluctuation in the ‘effect’¹³.

The mlVAR model, however, faces several challenges that need to be addressed¹⁴. Firstly, some assumptions connected to the model are frequently violated. For instance, the assumption of stationarity, assuming that the mean, variance, and autocorrelation of variables remain constant over time¹⁵, is violated in developmental and treatment processes¹⁶. In previous research using mlVAR in the analysis of EMA data, stationarity has been assumed due to the relatively short time span in which EMA data are captured¹³. However, this is not always realistic, as developments and changes can occur in the individual in short time periods. Furthermore, it precludes the study of networks where change is expected, for instance during an intervention¹⁷. Secondly, the undirected and between-subjects mlVAR approaches rely on partial correlations, and the directed mlVAR approach relies on multivariate regression, meaning that associations between two mood items are dependent on their relationships with all other mood items. This means that when all relevant variables are included and statistical assumptions are met, causality can be inferred from the connections in the directed network. However, in psychological systems a common occurrence is the presence of colliders—items that are influenced by two or more statistically unrelated nodes. Conditioning on the relationship with these colliders in the partial correlations network introduces a negative statistical dependency between the items influencing the collider, which can show up as a negative connection in the network when no such relationship exists in the data. Therefore, in psychological systems where colliders are present, the mlVAR approach might introduce spurious connections or distort the strength of positive connections in the network, leading to misinterpretations regarding the importance of certain nodes and edges¹⁸. Previous research has indeed demonstrated instances where mlVAR models produced such spurious connections in the network, which might be the consequence of such colliders^{18,19}. Additionally, items in the network may have construct overlap and can be highly intercorrelated, such that relationships among items can change noticeably if such an item is added or removed from the multivariate mlVAR network.

Dynamic Time Warping (DTW) offers a promising alternative for examining the relationships between items in a dynamic network. DTW is a statistical method that identifies patterns in time-series data by optimizing alignment, even when the co-variation between symptoms is not simultaneous or linear²⁰. This technique quantifies the shape-based similarity between variable trajectories over time, with the possibility of incorporating time lags, resulting in a DTW distance measure. Where mlVAR aims to identify causal connections, the DTW technique can be applied to study co-occurrence and synchrony between variables, with the possibility of identifying lagged and non-linear relationships. DTW has been applied to studying time-series data across various fields^{21–23}. In the field of psychiatry, DTW has recently shown success in estimating symptom networks of depression⁸, and examining changes in depression symptom networks after electroconvulsive therapy²⁴. It has also recently been used to map symptoms of post-traumatic stress²⁵, bipolar disorder^{26,27}, and eating disorder psychopathology²⁸.

Unlike mlVAR, DTW is a non-linear technique utilized for comparing and aligning time-series data, making it less susceptible to the effects of non-stationarity. Moreover, in DTW the calculation of the relationships between items is independent of their relationships with other items in the network. This means that the choice of variables included in the network does not influence the DTW distance between other variables. Additionally, using the DTW algorithm makes it easier to incorporate time-varying lags in the model. This allows for the inclusion of lag-0, lag-1, and lag-2 co-variation within the same model²⁹, unlike mlVAR, which requires advanced script extensions or custom equations³⁰. However, DTW also has its potential downsides, as it cannot capture multivariate associations or autocorrelation. Therefore, under ideal conditions where assumptions are met, Granger causality may be more reliably inferred from mlVAR networks. However, the mlVAR network is more susceptible to unmeasured confounding, as this can distort the mlVAR results but not the DTW distance calculations.

The objective of this study was to assess the advantages and disadvantages of the novel DTW approach and the conventional mlVAR approach by applying both to creating mood item networks based on the same dataset. Additionally, we aimed to investigate whether the techniques were susceptible to spurious or distorted

connections due to the presence of colliders and violations of the statistical assumptions of mIVAR that are common in psychological research, by repeating the analyses on simulated datasets. A script for these analyses was made available on OSF and included in the supplementary files. We hypothesized that the DTW algorithm would yield more reliable directed networks when collider variables were present, due to the potential vulnerability of the mIVAR temporal network to spurious and distorted associations under such circumstances.

Methods

Participants and procedure

The Netherlands Study of Depression and Anxiety (NESDA)³¹ is a multi-center cohort study that has included a wide range of psychological and biological measures, collected in seven waves thus far. The baseline data, known as the first wave (T0) were gathered from 2004 to 2007 ($n=2981$). The Assessment and Actigraphy module (NESDA-EMAA) module was offered to participants in the sixth wave (T5, collected between 2010 and 2013; $n=2069$ original participants and 367 siblings from the newly added sibling cohort). Participants from the sibling cohort without a current or past diagnosis of depressive and/or anxiety disorder or other severe psychiatric disorder were also eligible³². In the current study, we included participants of NESDA-EMAA with at least 35 completed time points in the final sample. The study was approved by the Medical Ethical Committee of the VUmc (reference number 2003/183) and written informed consent was obtained from all respondents. Research was conducted in accordance with the Declaration of Helsinki.

Measures

Demographics

Basic demographics and clinical characteristics were collected at each NESDA wave, including gender, age, and diagnostic status based on the Composite International Diagnostic Interview³³.

Momentary affective states

The NESDA-EMAA module consisted of 30 items. For the current study, we focused on items 1 through 20. Item 1–18 assessed affect and cognition, and 19 (“*I feel tired*”) and 20 (“*I experience physical discomfort*”) aimed to capture participants’ physical condition. The module included ten negative mood items, such as “*I feel irritated*” and eight positive mood items, such as “*I feel cheerful*”. All items were rated on a seven-point Likert scale, ranging from 1 (“*Not at all*”) to 7 (“*Very*”). Participants were prompted to complete the items five times a day for a duration of 2 weeks. This resulted in a potential total of up to 70 completions per participant.

Statistical analysis

We constructed distinct undirected and directed networks of mood items using the DTW and mIVAR algorithms. To further examine the (dis)similarities in outcomes between both techniques, we replicated the directed network analyses on a simulated dataset. This simulated dataset was intentionally designed to include a small number of predetermined associations between variables, along with random error/noise. This setup allowed us to investigate the origins of the potential differences that emerged.

Prior to conducting the analyses, all mood item scores were standardized at the group level. This standardization ensured that the results were based on the relative changes in mood item scores over time, enabling a meaningful comparison of trajectories even when scores fluctuated with different mean severity levels.

Sociodemographic and clinical variables at baseline are summarized as means and standard deviations (*SD*) or percentages, as appropriate. Additionally, an unadjusted density plot of all the 24,422 assessments was generated to visualize the relationships between item scores in the dataset, using the R ‘*corrplot*’ package (version 0.92), which provided insight into the crude interdependencies.

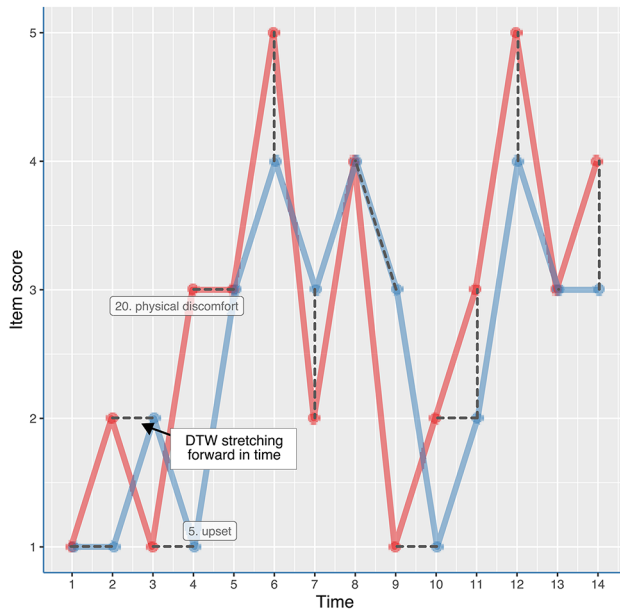
Undirected analyses The contemporaneous and between-subjects Multi-level Vector AutoRegression models For the mIVAR analysis, we utilized the “mIVAR” package in R^{34,35}. This package contains an algorithm that estimates the predictive relationships between each variable and all other included variables at the subsequent time point within a nested model. As the EMA data were collected in regular time intervals during the day, but not collected during the night, we instructed the model to account for these timing discrepancies by not regressing the first measurement point of a day over the last measurement point of the previous day. Importantly, the mIVAR model separates within-person from between-person effects to account for within-subject dependence of data points and non-ergodicity. Analysis therefore yields two separate undirected network models. The “contemporaneous” (within-subject) network estimates cross-sectional within-subject associations between mood items, while controlling for temporal relationships and all other items. The “between-subjects” network estimates associations between mean scores of the items across timepoints and across subjects³⁵. For a more elaborate description of the mIVAR model, see Bringmann et al.¹⁰ and Epskamp et al.³⁵

Undirected Dynamic Time Warping

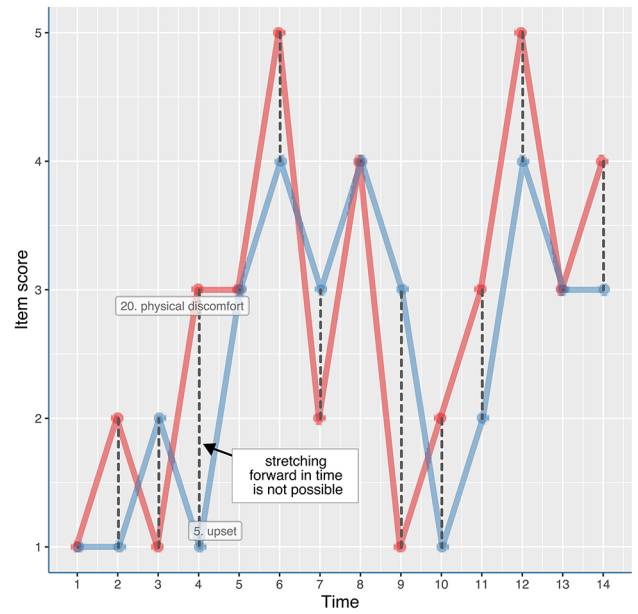
The DTW networks were constructed based on the calculation of DTW distances between each pair of EMA mood items, which reflect the tendency of the items to co-vary. Each mood item’s score trajectory over time was treated as an individual time series. Unlike simple point-to-point Euclidian comparisons commonly used in longitudinal studies, the DTW technique offered a more flexible alignment by considering both the cross-sectional distance (t , *lag-0*) and the distance with a lagged point in time ($t-1$ or $t+1$). This flexibility allowed for shifts and deformations in the time axis, enabling a comprehensive analysis of dynamic patterns. For a detailed explanation and visualization of the DTW technique, see Van der Does et al.²⁵.

The DTW distance calculation involved creating a symmetric cost matrix for each of the up to 70×70 item combinations. To mitigate the potential disruptive effects of starting and endpoint mismatches, we interpolated five values between time points to ensure a more accurate distance calculation. The undirected DTW algorithm

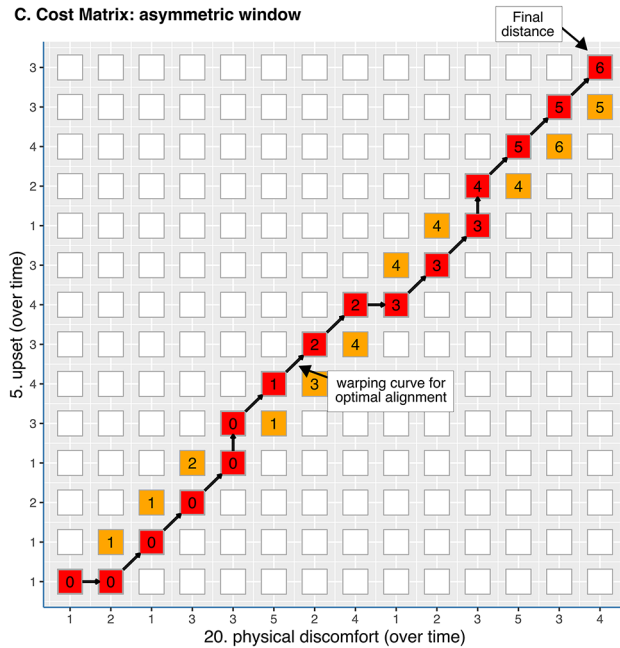
A. Directed: Item 20 predicting Item 5



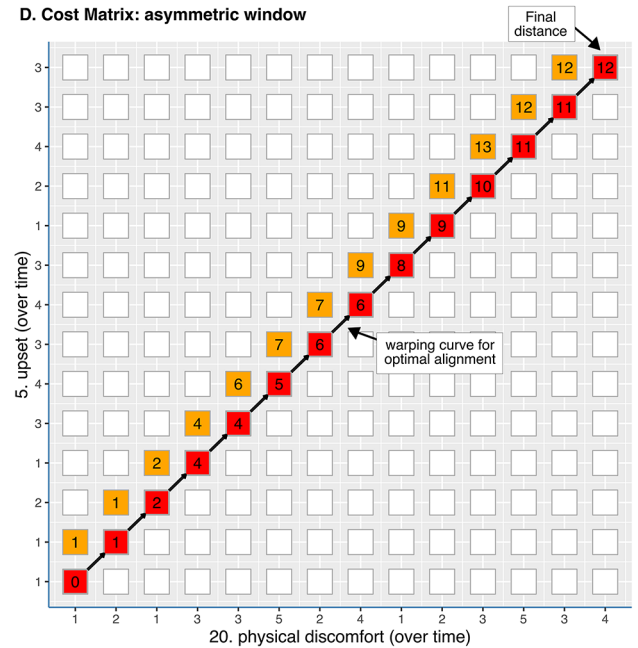
B. Directed: Item 5 predicting Item 20



C. Cost Matrix: asymmetric window



D. Cost Matrix: asymmetric window



D. Calculation of the directed distances

$$\text{Distance}_{\text{Item20 to item 5}} = \frac{\text{distance_from} - \text{distance_to}}{\text{mean}(\text{distance_from}, \text{distance_to})} = \frac{12 - 6}{12 + 6} = \frac{6}{18} = 0.33$$

$$\text{Distance}_{\text{Item 5 to item 20}} = \frac{\text{distance_to} - \text{distance_from}}{\text{mean}(\text{distance_from}, \text{distance_to})} = \frac{6 - 12}{6 + 12} = \frac{-6}{18} = -0.33$$

E. Step pattern: "symmetric2"

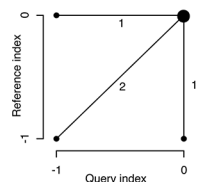
Recursion:

$$g[i, j] = \min ($$

$$g[i-1, j-1] + 2 * d[i, j],$$

$$g[i, j-1] + d[i, j],$$

$$g[i-1, j] + d[i, j])$$



was then applied to identify the optimal path that minimized the total accumulated distance, resulting in the DTW distance. The algorithm was restricted to matching time points on the first variable to time points on the second variable that were simultaneous (t) one step ahead ($t + 1$) and one step backwards ($t - 1$) through a 'Sakoe-Chiba' window band of 1. The Sakoe-Chiba window constrains DTW by limiting how far time points can shift. In undirected DTW, it allows time points to stretch both forward and backward, whereas in directed DTW, it restricts stretching to only forward shifts. Other methods for constraining DTW include the asymmetric Itakura parallelogram and unconstrained DTW, which lacks any restrictions³⁶. The Sakoe-Chiba window maintains consistent constraints throughout the alignment process, making it a preferred choice in many scenarios. It is often considered superior to the Itakura parallelogram³⁷, and unconstrained DTW can lead to overfitting by producing unnatural alignments³⁸. The default 'symmetric2' step pattern was used to match the sequences, giving both time series equal priority and allowing for diagonal steps through the cost matrix, though prioritizing

◀ **Fig. 1.** The calculation of the directed DTW distance between the time-series data of two symptom items. The directed DTW distance between items “5” and “20, over 14 time points, is calculated both for the scenario where fluctuations in item 20 tend to precede fluctuations in item 5 (panel A) and the scenario where fluctuations in item 5 tend to precede fluctuations in item 20 (panel B). The first step in DTW is creating local cost matrices (LCM), which in this case have 14×14 dimensions (panels C and D). In the second step, the DTW algorithm finds the path that optimizes the alignment between the two time series by iteratively stepping through the LCM, starting at the lower left corner and finishing at the upper right corner, while aggregating the total distance (i.e., “cost”). At each step, the algorithm takes the step in the direction in which the cost increases the least under the chosen constraint. The constraint was the Sakoe–Chiba window of size one, meaning one time-point after the current assessment. The way in which the algorithm traverses through the LCM is dictated by the chosen step pattern, in this case the default “symmetric2” step pattern (F). Panels (E) and (F) explain the calculations of DTW distances in both directions for the symptom pair, yielding 6 and 12 as their respective distances. On the basis of this lower directed DTW distance we can conclude that fluctuations in item 20 tend to precede fluctuations in item 5, more than the other way around.

horizontal or vertical steps³⁶. The resulting distance was normalized based on the number of assessments within that individual. A lower distance indicated more similar dynamics between two time-series, signaling a higher similarity of fluctuations over time.

For each pair of variables, two distances were computed: one representing the distance from mood item A to mood item B, and another representing the distance from mood item A to the inverse of mood item B. We then selected the smallest absolute value among these two distances. This approach accounted for the presence of mood items either with a positive and negative valence in the dataset (see Supplementary Fig. 1). This analysis considered the possibility of both negative and positive feedback loops existing in the network, aiming to capture the complex interplay and dynamics among the variables. This process resulted in 376 symmetrical distance matrices, each representing an individual participant. These were combined to create a single overarching distance matrix (see Supplementary Table 1). We then fit a linear model to evaluate how strongly the two variables under study influenced the DTW distance, adjusted for the variance in item-scores within each participant.

The resulting undirected network plot illustrated the dynamic interactions between the 20 mood items. Edges were included if the distance between two items was significantly smaller ($p < 0.05$) than the average of all remaining distances, after adjusting for the variance of the scores of both items over time for each of the participants. This adjustment accounted for the inflated clustering of scores that remained zero or constant throughout follow-up. We performed a hierarchical cluster analysis using the “Ward. D2” clustering method to identify clusters of mood items exhibiting similar dynamic patterns, reflected in the color-coding of nodes in the network. The primary objective was to group items within the same dimension while maximizing dissimilarity between items from different dimensions. The thickness and color intensity of the edges in the plot corresponded to the strength of association between two items, while thinner and lighter-colored edges represented weaker associations.

Directed analyses The temporal Multi-Level Vector AutoRegression model

The “mlVAR” package can also yield a time series network, the “temporal” network. When using this package, one has to decide what lag to include in the model. Commonly, a lag-1 model is used, meaning that the model assesses to what extent the score on an item is predicted by other items and itself one time-lag in the past. The autoregressive effects have been removed from this model, to increase legibility of the plot, and comparability with the directed DTW network. For a figure that includes the autoregressive effect, see Supplementary Fig. 2.

Directed Dynamic Time Warping

The directed analyses of variables focused on assessing the direction of fluctuations in items, or what changes in item scores tended to precede changes in other item scores. We utilized the same DTW algorithm as in the undirected analyses but with an asymmetric time window, as a variation of the Sakoe–Chiba band. This band restricted the dynamic alignment between items to the current (t) and the next time point ($t + 1$) (see Fig. 1).

To determine the direction of the relationship between mood item A and mood item B, we calculated the distance from item A to B (Fig. 1, panels A, C) and the distance from mood item B to mood item A [Fig. 1, panels B, D]. The difference was divided by sum of both distances which yielded the relative distances from B to A and from A to B. This procedure was repeated for the inverse of mood item B, again to account for the presence of both positive and negative valence of mood items in the dataset [see Supplementary Fig. 1]. This resulted in two distance matrices per participant, one for the positive associations between items and one for the inverse associations between items. We calculated the group-level directed network plot by testing whether the positive or inverse time-lag association was significantly different from zero. In the directed network plot, statistically significant directed edges were represented by arrows, with the arrowhead pointing in the temporal direction. If the D was not significantly different from 0, no arrow was included in the plot. A positive value of D indicated a positive feedback effect, represented by a green arrow from mood item A to mood item B. Conversely, a negative value of D indicated an inverse feedback effect (e.g., when A increases B tends to decrease one timepoint later), and a red arrow from A to B was included. To assess the centrality of each item, we utilized standardized measures of temporal lead and temporal lag. Temporal lead reflects the number and strength of outgoing arrows, with significant temporal lead indicating that fluctuations in this item tended to precede that of other items. Conversely, the temporal lag reflected the number and strengths of incoming arrows, with significant temporal lag indicating that changes in the item followed changes in other items.

Simulated datasets We conducted simulations using nine datasets, each consisting of 200 participants and 200 measurement points, to identify any spurious, missing, or distorted connections introduced by the statistical models. These simulated datasets were designed to have strong predetermined associations, and included random noise with a range of $[-0.2$ to $0.2]$ to approximate real-world measurements. No repetitions were included in the analyses, as these were not expected to yield different results. The aim of our simulation analysis was to evaluate whether the mIVAR and DTW algorithms accurately detected the known associations explicitly embedded in the data, rather than to infer or assess the underlying relationships within the dataset. The datasets each included two, three or four variables. Additionally, we manipulated the presence and number of variables with changes preceding variations in collider variables, allowing us to examine the potential bias introduced by colliders. Finally, we varied the lag in the co-variation of variables, including scenarios with a perfect lag-1, partial lag-1 (50% perfect lag-1 and 50% lag-0), or partial lag-2. By systematically manipulating the number of variables, the presence of colliders, and the lag in co-variation, we aimed to gain insights into any differences between the directed networks obtained from the mIVAR and DTW algorithms. The sample code is given in the supplementary file and on online Open Science Framework via https://osf.io/sample_code.

All statistical analyses were performed using R version 4.2.2³⁹, using the packages “dtw” (version 1.23–1), “parallelDist” (version 0.2.6), “mIVAR” (version 0.5) and “qgraph” (version 1.9.3).

Results

Sample characteristics

The final study sample consisted of 376 participants; 347 original participants and 29 from the sibling cohort. The participants' mean age was 49.3 ($SD=12.7$), and 242 were female (64.4%). 99 participants (26.3%) had a current diagnosis of anxiety and/or depression. Specifically, 29 (7.7%) participants were currently experiencing depression, 38 (10.1%) had a current anxiety disorder, and 32 (8.5%) had a current comorbid anxiety and depression diagnosis. The mean number of assessments per participant was 65.0 ($SD=5.0$).

Undirected analyses

In the undirected DTW network (Fig. 2, panel A), we observed stronger positive connections among the items in the Positive Affect cluster compared to the mIVAR networks. Additionally, there were strong relationships between some negative items, for instance the relationships of the item Irritated with the items Nervous and Down, and the relationship between the items Upset and Anxious. For the undirected DTW distance matrix, see Supplementary Table 1.

The mIVAR contemporaneous network (Fig. 2, panel B) revealed a higher number of stronger negative edges, particularly between the Positive Affect cluster and both the Negative Affect and Somatic clusters. This indicates that, within the same measurement period, stronger negative moods were associated with weaker positive mood items, as one might expect. Additionally, the strongest connections in this network were the negative associations between Tired and Energetic, Tense and Relaxed, and Tense and Calm, reflecting mood states that are intuitively indeed unlikely to occur simultaneously.

The mIVAR between-subjects network revealed several strong positive connections that were either absent or much weaker in the DTW network. These connections included associations between Physical Discomfort and Tired, Worried and Down, and Anxious and Nervous.

Directed Analyses

The mIVAR network displayed a larger total number of connections compared to the DTW network (see Fig. 3, panels A and B). However, some of the connections that were particularly important in the DTW network were weak or missing in the mIVAR network. For instance, the edges between Energetic and Satisfied and Enthusiastic and Satisfied were prominently present in the DTW network, but appeared less strong in the mIVAR network. Furthermore, some of the prominent edges of the mIVAR network were not present in the DTW network, like the edges between Physical Discomfort and Tired, and Worried and Nervous. Notably, we removed autoregressive effects from the model in the mIVAR network in Fig. 3, to enhance visual comparability to the directed DTW

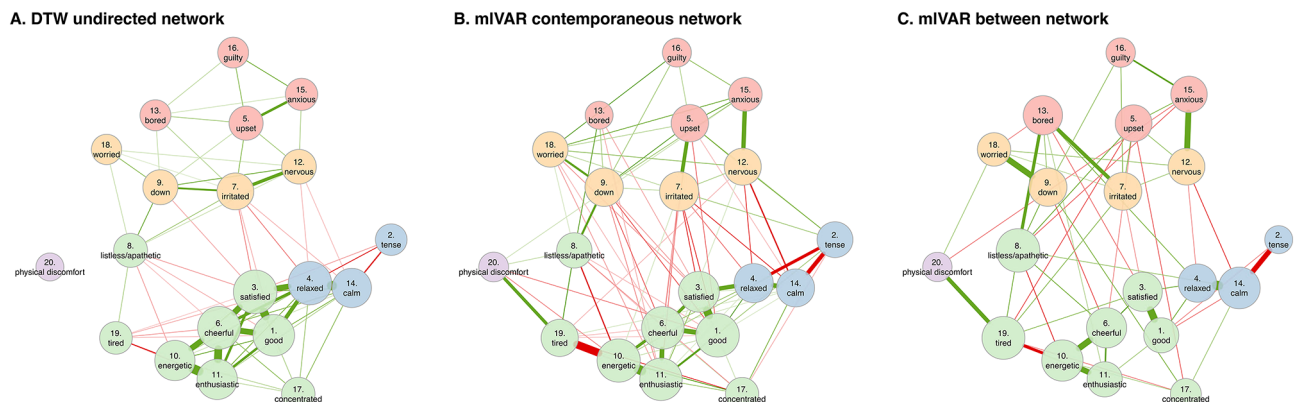
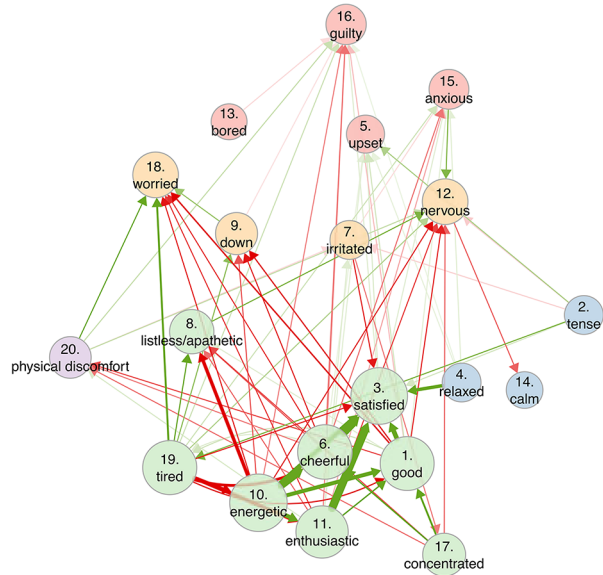
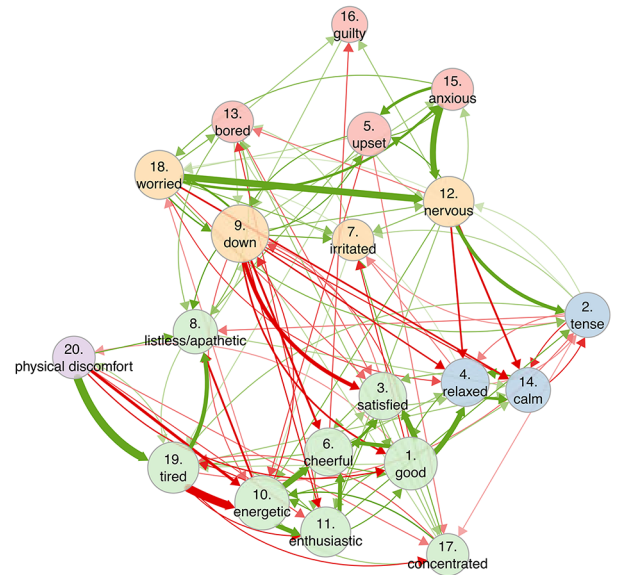


Fig. 2. The undirected DTW network, and the mIVAR contemporaneous and between-subjects networks.

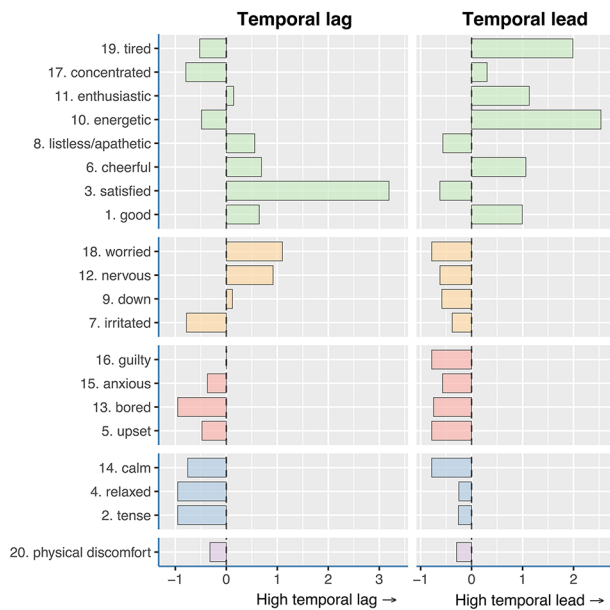
A. DTW directed network



B. mIVAR temporal network



C. DTW temporal lead and lag



D. mIVAR in- and out-strength

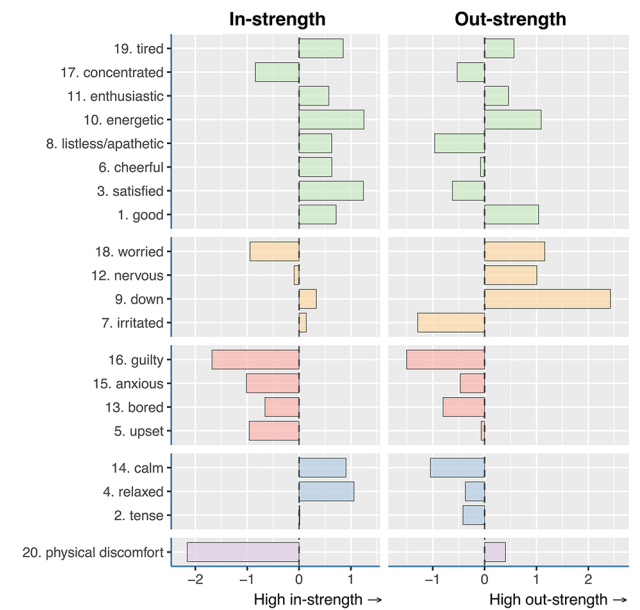


Fig. 3. The directed DTW and temporal mIVAR mood item networks. The DTW networks’ temporal lag and temporal lead, and the mIVAR network’s in- and out-strength centrality are displayed in corresponding bar graphs. Autoregressive effects were removed from the mIVAR network to facilitate comparison.

network. We have included an mIVAR network that includes these autoregressive effects in Supplementary Fig. 2. In this figure (Supplementary Fig. 2, panel C), the connections between mood items within the mIVAR network appear less strong than when the autoregressive effects are omitted, but the structure of the network is very similar (Supplementary Fig. 2, panel B). Therefore, the differences with the DTW network (Supplementary Fig. 2, panel A) cannot be explained by the removal of the autoregressive effects. This is to be expected, as the algorithms measure different types of relationships between mood items.

Centrality profiles of the mood items differed for the DTW and mIVAR networks (see Fig. 3, panels C and D). In the DTW network, we observed the highest temporal lag for the items “I feel satisfied” from the positive affect cluster and “I feel worried” and “I feel nervous” from the negatively valenced items. This indicates that fluctuations in these variables tended to follow fluctuations in other variables, for instance suggesting that changes in feelings of satisfaction may, to some extent, be influenced by the dynamics of other variables. The items for which we observed the strongest temporal lead, indicating that fluctuations in these variables tended to



Fig. 4. Comparing the associations found between variables using the mIVAR and DTW algorithms and simulated data. Datasets were simulated with variations in the number of variables, the presence of colliders, time lag, and completeness of the time lag.

precede fluctuations in other variables, were “*I feel tired*”, and “*I feel energetic*”. These nodes may exert influence over other nodes within the DTW network, although this potential interdependence has not been adjusted for confounding factors. For the directed DTW distance matrix, see Supplementary Table 2.

In the mIVAR temporal network, the items with the highest in-strength were “*I feel energetic*” and “*I feel satisfied*”. As was the case in the DTW network, this indicates that satisfaction may be influenced by other mood items. The highest out-strength was observed in the item “*I feel down*”, reflecting the highest direct predictive value when relationships with the other variables included in the network were controlled for. When interpreted through Granger causality⁴⁰, this item may play an important role in driving changes within the network.

Simulated datasets

In the simulated dataset, the results of the mIVAR and DTW analyses were identical when a complete lag-1 was simulated. This means that all fluctuations in items consistently preceded or followed fluctuations in other items with a time lag of exactly one time point. However, when a partial lag was introduced, meaning that the co-variation was not strictly lag-1 but also included instantaneous (lag-0) effects, spurious connections emerged in the mIVAR network when collider variables were present. These spurious connections indicated associations between items that were not actually present in the underlying data. Furthermore, when the algorithms were tested with a lag-window of 2, the mIVAR network no longer reflected the lag-1 associations within the network. On the other hand, using the DTW algorithm allowed for incorporating both the lag-1 and lag-2 time windows in a single analysis, still revealing the lag-1 relationship that was present in the data.

The results demonstrate that the DTW algorithm correctly identified the associations that were present in the simulated data, regardless of the lag in item fluctuations. In contrast, the mIVAR models produced spurious or distorted connections when the lag in item fluctuations was not complete (see Fig. 4). Furthermore, the standard mIVAR algorithm was not able to simultaneously reflect multiple time lag effects without mathematical extensions of the model.

Discussion

In this study, we aimed to compare the advantages and disadvantages of the innovative DTW technique with the more conventional mIVAR method for constructing undirected and directed mood item networks. Direct comparisons between the networks are challenging, as the mIVAR and DTW algorithms capture different underlying effects, and differences are thus to be expected. The undirected mIVAR networks aim to capture

the associations between variables that occur simultaneously (contemporaneous), and associations across individuals (between-subjects). Visual inspection of the undirected DTW network and the contemporaneous mIVAR network shows that basing conclusions about connections within the networks on shape-based alignment of the time-series data, as in DTW, would lead to very different conclusions than when using the mIVAR contemporaneous network, which captures same-time associations, or the mIVAR between-subjects network, that aims to capture stable properties across subjects. Notably, even though in mIVAR networks the connections are partial correlations and could reasonably be expected to be sparser due to controlling for within-network confounding⁴¹, the DTW connections appear overall less strong. When looking at the undirected DTW network, there seems to be stronger time-lagged co-fluctuation among positive mood items, and less association between positive and negative mood items compared to the contemporaneous network. The disparities within the directed networks based on DTW and mIVAR raise more significant concerns, as interpretations regarding the directional influence and importance of symptoms or other items that would likely emerge from these networks stand in substantial contrast. Subsequent analyses conducted on simulated datasets unveiled the mIVAR technique's susceptibility to introducing spurious and distorted connections within networks of interrelated items, especially those with complex interrelationships.

Although the connections present in the mIVAR and DTW networks could not be directly compared as they reflect different types of relationships, in certain conditions often present in real-world psychological data the mIVAR network may be more prone to reflecting spurious connections. In the evolving field of network analysis in psychopathology, there is no current consensus on how the symptom networks based on EMA data are best modelled. The use of diverse statistical techniques may yield vastly different interpretations of the importance of symptom nodes⁴². Therefore, clinical decisions based on these networks may be heavily influenced by the statistical decisions made in the construction of the algorithm. The directed DTW and temporal mIVAR networks exhibited numerous dissimilarities, apparent in their graphical representations and in- and out-strength profiles. For instance, the mIVAR network suggested “*I feel down*” as the most influential item due to its high out-strength. Interpreting this finding through Granger causality⁴⁰ implies that fluctuations in this symptom may have a relatively large effect on other symptoms in the network, and therefore may be interpreted as an important target for treatment⁴². However, in the directed DTW network this item had low temporal lead, and energy-related items “*I feel tired*” and “*I feel energetic*” had the highest temporal lead. Therefore, the algorithm used to analyze the data and generate the network could result in significant differences in interpretation of the importance of the mood items. An important caveat of the literature on mIVAR is that it relies on Granger causality, which offers limited evidence for true causality⁴³. Also, this form of causality cannot be inferred from DTW analyses due to the lack of conditioning on other relationships. This limitation could be addressed in the future with techniques like Convergent Cross-Mapping⁴⁴, which aims to detect causality in complex systems by assessing whether one time series can reliably predict another.

To evaluate algorithm performance, we conducted analyses on simulated datasets, systematically varying the number of variables, colliders, and time lags in symptom covariation. Simulated data, in which the associations were predetermined, allowed us to reliably identify spurious or distorted connections. These simulations highlighted the mIVAR algorithm's vulnerability to generating spurious or distorted connections when incomplete lags and colliders were present. This is an important finding, as psychological processes often involve complex relationships and items with construct overlap. Therefore, varying time lags and colliders are likely to occur. Drawing erroneous conclusions about the directionality and importance of symptom effects may be consequential in the choice of targets for treatment. On the basis of our results, we urge caution in the interpretation of temporal mIVAR networks in case of nodes with significant construct overlap and high intercorrelation. The directed DTW network could offer a promising alternative, although this technique has some significant downsides that we will also discuss in this paper.

A key distinction between the undirected DTW and mIVAR techniques lies in their approach to handling potential non-ergodicity within the data. The DTW technique initially computes within-person dynamics, and subsequently aggregates these in the network^{8,24–28}. Conversely, the mIVAR method generates two separate networks – a contemporaneous (within-measurement) network and a between-subjects network¹². This makes the networks based on the different algorithms difficult to directly compare. Another substantial difference between the approaches is that undirected mIVAR is based on partial correlations and the directed mIVAR on multivariate regression, while the DTW algorithm is a bivariate technique. In the mIVAR network, all connections between symptoms are conditional on all other connections within the network³⁵. This means that the network structure changes depending on the variables measured and decisions made by the researchers on inclusion or omission of variables within the network. In contrast, because of the bivariate nature of the DTW algorithm connections in this network are independent of the absence or presence of other variables. However, this also presents some potential downsides, as the DTW algorithm's ability to capture complex relationships between multiple variables in a network may be hampered. For example, if A leads to B and B leads to C, there will also be an edge between A and C using DTW, which may not be present in mIVAR. Although this makes the network robust against variable selection bias, it is vulnerable to associations between nodes that could be explained by associations with other variables in the network. Furthermore, the aggregate temporal lead and lag in the directed DTW analysis does not provide meaningful insights into the network. Since the relationships between variables are not conditioned on one another, the total lead or lag of a network increases as more variables are included.

A limitation of the calculation of the significance of edges in undirected DTW networks is that it is dependent on the average value of the other edges, meaning that introducing or removing variables can affect this average. Furthermore, in very unusual cases where the time lags between variables are uniform but non-zero, the network may yield no significant connections.

A limitation of the directed DTW technique is that it is not able to capture autoregressive effects, which reflect how the value of a mood item predicts its own future value¹². Namely, the DTW distance between the time-series measurement of a variable and itself is always zero. In the current study, we also removed autoregressive effects from the temporal mIVAR network to enhance comparability with the directed DTW network. Including autoregressive effects may provide valuable insights into the dynamics of the network, as the persistence of certain symptoms over time may point to their clinical importance⁴². To address this, we included a *Supplementary Fig. 2* comparing the mIVAR networks with and without these autoregressive effects. This figure shows that the differences in the networks are not artifacts of removing these autoregressive effects from the mIVAR analysis.

We believe that this study makes a substantial contribution to the existing network literature, as it is the first to compare the novel DTW technique to the more conventional mIVAR method of modelling EMA networks. It included a large and varied sample of participants with current or remitted psychopathology and healthy controls, and a high number of measurements per participant. The EMA method used in this study is especially useful in the collection of data for personalized networks, as it is able to generate a large amount of data at a low cost to the patient or participant. The high number of measurement points facilitates mapping a nuanced image of short-term interplay between mood states. Nevertheless, the DTW technique may also outperform mIVAR in analyzing sparse panel data. The mIVAR approach is not suitable for sparse data as it is susceptible to Nickell's bias, meaning that when fewer than 20 to 60 measurement points are included, more spurious autoregressive connections could be introduced in the network¹². Therefore, the DTW technique could for instance provide significant advantages for the study of serious psychiatric disorders or geriatric psychiatry, as research participants may be unable to complete extensive numbers of assessments⁴⁵. Our use of simulated datasets to evaluate the DTW and mIVAR algorithms provides important insights for future research into understanding psychiatric disorders as complex dynamic systems. This study represents another step towards the clinical applicability of network theory in research, clinical practice, and personalized care.

Conclusion

We have demonstrated that the temporal mIVAR technique, while offering some theoretical advantages over the DTW algorithm, exhibits a vulnerability to introducing spurious and distorted connections in the network representation of psychological processes. The DTW network's resilience to colliders, varying time lags and variable selection bias and non-stationarity underscore its potential as a more stable alternative to mIVAR networks.

Data availability

Scientists can apply for data and samples through submitting a research proposal to the NESDA consortium (nesda@ggzingeest.nl). We have published the script of our simulation analysis online at https://osf.io/nxa2v/?view_only=b7a07edbbc9047ffbf8cd0d6f6274925.

Received: 29 July 2024; Accepted: 17 March 2025

Published online: 05 April 2025

References

- Borsboom, D. & Cramer, A. O. Network analysis: an integrative approach to the structure of psychopathology. *Ann. Rev. Clin. Psychol.* **9**, 91–121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608> (2013).
- Blanchard, M. A. & Heeren, A. Ongoing and future challenges of the network approach to psychopathology: From theoretical conjectures to clinical translations. In *Reference Module in Neuroscience and Biobehavioral Psychology*. (Elsevier, 2020). <https://doi.org/10.1016/B978-0-12-818697-8.00044-3>
- Fried, E. I., Flake, J. K. & Robinaugh, D. J. Revisiting the theoretical and methodological foundations of depression measurement. *Nat. Rev. Psychol.* **1**, 358–368. <https://doi.org/10.1038/s44159-022-00050-2> (2022).
- Armour, C., Fried, E. I. & Olf, M. Vol. 8, 1398003 (Taylor & Francis, 2017).
- Galatzer-Levy, I. R. & Bryant, R. A. 636,120 ways to have posttraumatic stress disorder. *Perspect. Psychol. Sci.* **8**, 651–662 (2013).
- Bringmann, L. F. & Eronen, M. I. Don't blame the model: Reconsidering the network approach to psychopathology. *Psychol. Rev.* **125**, 606 (2018).
- McNally, R. J. Can network analysis transform psychopathology?. *Behav. Res. Therapy* **86**, 95–104. <https://doi.org/10.1016/j.brat.2016.06.006> (2016).
- Hebbrecht, K. et al. Understanding personalized dynamics to inform precision medicine: a dynamic time warp analysis of 255 depressed inpatients. *BMC Med.* **18**, 1–15. <https://doi.org/10.1186/s12916-020-01867-5> (2020).
- Lutz, W. et al. Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: A methodological proof-of-concept study. *Sci. Rep.* **8**, 7819. <https://doi.org/10.1038/s41598-018-25953-0> (2018).
- Bringmann, L. F. et al. A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS One* **8**, e60188. <https://doi.org/10.1371/journal.pone.0060188> (2013).
- Schuurman, N. K. *Multilevel Autoregressive Modeling in Psychology: Snags and Solutions*. Utrecht University (2016).
- Jordan, D. G., Winer, E. S. & Salem, T. The current status of temporal network analysis for clinical science: Considerations as the paradigm shifts?. *J. Clin. Psychol.* **76**, 1591–1612 (2020).
- Epskamp, S. Discovering psychological dynamics. *Network psychometrics*, 85–114 (2017).
- Bringmann, L. F. The future of dynamic networks in research and clinical practice. *World Psychiatry: Off. J. World Psychiatr. Assoc. (WPA)* **23**, 288–289 (2024).
- Takano, K., Stefanovic, M., Rosenkranz, T. & Ehring, T. Clustering individuals on limited features of a vector autoregressive model. *Multivar. Behav. Res.* **56**, 768–786. <https://doi.org/10.1080/00273171.2020.1767532> (2021).
- Bringmann, L. F., Ferrer, E., Hamaker, E. L., Borsboom, D. & Tuerlinckx, F. Modeling nonstationary emotion dynamics in dyads using a time-varying vector-autoregressive model. *Multivar. Behav. Res.* **53**, 293–314. <https://doi.org/10.1080/00273171.2018.1439722> (2018).
- Bringmann, L. F. et al. Assessing temporal emotion dynamics using networks. *Assessment* **23**, 425–435 (2016).
- Ryan, O., Bringmann, L. F. & Schuurman, N. K. The challenge of generating causal hypotheses using network models. *Struct. Equ. Model. Multidiscip. J.* **29**, 953–970. <https://doi.org/10.1080/10705511.2022.2056039> (2022).

19. Greene, T., Gelkopf, M., Epskamp, S. & Fried, E. Dynamic networks of PTSD symptoms during conflict. *Psychological Medicine* **48**, 2409–2417. <https://doi.org/10.1017/S0033291718000351> (2018).
20. Keogh, E. J. & Pazzani, M. J. In *European Conference on Principles of Data Mining and Knowledge Discovery*, 1–11 (Springer).
21. Gavrilu, D. M. & Davis, L. S. In *International Workshop on Automatic Face-and Gesture-Recognition*, 272–277 (Citeseer).
22. Amin, T. B. & Mahmood, I. In *2008 2nd International Conference on Advances in Space Technologies*, 74–79 (IEEE).
23. Aach, J. & Church, G. M. Aligning gene expression time series with time warping algorithms. *Bioinformatics* **17**, 495–508. <https://doi.org/10.1093/bioinformatics/17.6.495> (2001).
24. Booi, M. M. et al. Dynamic time warp analysis of individual symptom trajectories in depressed patients treated with electroconvulsive therapy. *J. Affect. Disord.* **293**, 435–443. <https://doi.org/10.1016/j.jad.2021.06.068> (2021).
25. Van der Does, F. H. et al. PTSD symptom dynamics after the great east japan earthquake: mapping the temporal structure using dynamic time warping. *Eur. J. Psychotraumatol.* **14**, 2241732. <https://doi.org/10.1080/20008066.2023.2241732> (2023).
26. Qian, Y., Solano, M. J. & Kreindler, D. Grouping of mood symptoms by time series dynamics. *J. Affect. Disord.* **309**, 186–192. <https://doi.org/10.1016/j.jad.2022.04.117> (2022).
27. Mesbah, R. et al. Dynamic time warp analysis of individual symptom trajectories in individuals with bipolar disorder. *Bipolar Disord.* <https://doi.org/10.1111/bdi.13340> (2023).
28. Slof-Op't Landt, M. C., Dingemans, A. E. & Giltay, E. J. Eating disorder psychopathology dimensions based on individual co-occurrence patterns of symptoms over time: A dynamic time warp analysis in a large naturalistic patient cohort. *Eat. Weight Disord.-Stud. Anorex. Bulimia Obes.* **27**, 3649–3663. <https://doi.org/10.1007/s40519-022-01504-5> (2022).
29. Gaskell, P., McGroarty, F. & Tiropanis, T. Signal diffusion mapping: optimal forecasting with time-varying lags. *J. Forecast.* **35**, 70–85. <https://doi.org/10.1002/for.2365> (2016).
30. Haslbeck, J. & Waldorp, L. J. mgm: Estimating time-varying mixed graphical models in high-dimensional data. arXiv preprint <https://doi.org/10.48550/arXiv.1510.06871> (2015).
31. Penninx, B. W. et al. The Netherlands study of depression and anxiety (NESDA): Rationale, objectives and methods. *Int. J. Methods Psychiatr. Res.* **17**, 121–140 (2008).
32. Van Genugten, C. R. et al. Experienced burden of and adherence to smartphone-based ecological momentary assessment in persons with affective disorders. *J. Clin. Med.* **9**, 322. <https://doi.org/10.3390/jcm9020322> (2020).
33. Organization, W. H. The Composite International Diagnostic Interview (1997).
34. Epskamp, S. Psychometric network models from time-series and panel data. *Psychometrika* **85**, 206–231. <https://doi.org/10.1007/s11336-020-09697-3> (2020).
35. Epskamp, S., Waldorp, L. J., Möttus, R. & Borsboom, D. The Gaussian graphical model in cross-sectional and time-series data. *Multivar. Behav. Res.* **53**, 453–480 (2018).
36. Giorgino, T. Computing and visualizing dynamic time warping alignments in R: the dtw package. *J. Stat. Softw.* **31**, 1–24 (2009).
37. Geler, Z., Kurbalija, V., Ivanović, M., Radovanović, M. & Dai, W. In *2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*. 1–6 (IEEE).
38. Zhang, Z. et al. Dynamic time warping under limited warping path length. *Inf. Sci.* **393**, 91–107 (2017).
39. Team, R. C. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. (2022).
40. Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econom.: J. Econom. Soc.* **37**, 424–438 (1969).
41. Borsboom, D. et al. Network analysis of multivariate data in psychological science. *Nat. Rev. Methods Primers* **1**, 58 (2021).
42. Bastiaansen, J. A. et al. Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *J. Psychosom. Res.* **137**, 110211 (2020).
43. Lusch, B., Maia, P. D. & Kutz, J. N. Inferring connectivity in networked dynamical systems: Challenges using Granger causality. *Phys. Rev. E* **94**, 032220 (2016).
44. Sugihara, G. et al. Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).
45. Van Zelst, D. C. et al. Network structure of time-varying depressive symptoms through dynamic time warp analysis in late-life depression. *Int. J. Geriatr. Psychiatry* <https://doi.org/10.1002/gps.5787> (2022).

Author contributions

Erik Giltay, Wessel van Eeden and Floor van der Does conceived the study. Erik Giltay and Floor van der Does performed the statistical analyses. Harriëtte Riese and Laura Bringmann consulted on the statistical analyses. Floor van der Does drafted the analysis plan and the manuscript. Erik Giltay supervised the project. All authors discussed the results and contributed to the final manuscript.

Funding

The infrastructure for the NESDA study (www.nesda.nl) has been funded through the Geestkracht program of the Netherlands Organisation for Health Research and Development (ZonMw, grant number 10-000-1002) and by participating universities and mental health care organizations (Amsterdam University Medical Centers (location VUmc), GGZ inGeest, Leiden University Medical Center, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, Rob Giel Onderzoekcentrum).

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval and consent to participate

The study was approved by the Medical Ethical Committee of the VUmc (reference number 2003/183) and written informed consent was obtained from all respondents.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-94782-9>.

Correspondence and requests for materials should be addressed to E.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025