bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The identification of temporal communities through trajectory clustering correlates with single-trial behavioural fluctuations in neuroimaging data

William Hedley Thompson[1,2], Jessey Wright[1,3], James M. Shine[4], & Russell A. Poldrack[1]

1. Department of Psychology, Stanford University, USA.
2. Department of Clinical Neuroscience, Karolinska Institutet, Sweden
3. Department of Philosophy, Stanford University, USA.
4. Brain and Mind Centre, The University of Sydney, NSW, Australia

Abstract

Interacting sets of nodes and fluctuations in their interaction are important properties of a dynamic network system. In some cases the edges reflecting these interactions are directly quantifiable from the data collected, however in many cases (such as functional magnetic resonance imaging (fMRI) data) the edges must be inferred from statistical relations between the nodes. Here we present a new method, called temporal communities by trajectory clustering (TCTC), that derives time-varying communities directly from time series data collected from the nodes in a network. First, we verify TCTC on resting and task fMRI data by showing that time-averaged results correspond with expected static connectivity results. We then show that the time-varying communities correlate and predict single-trial behaviour. This new perspective on temporal community detection of node-collected data identifies robust communities revealing ongoing spatial-temporal community configurations during task performance.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Many empirical phenomena can be mathematically described as networks, and an important property of networks is the presence of community structure. Communities are sets of nodes are more strongly interconnected with one another compared to the rest of the network (Fortunato & Hric, 2016; Newman, 2010). When collecting network data, information is sampled from the different nodes or edges. Edge-collected data, such as the number of emails sent between people, can seamlessly translate into a network representation. In contrast, node-collected data requires that edges be inferred based on a statistical relationship between nodes. This is typical of most non-invasive neuroimaging techniques, where recorded brain regions have their connectivity inferred from statistical relationships between the nodes' time-series (e.g. using Pearson's correlation). Only after this inference step can different network properties, such as modularity and nodal participation, be calculated.

Whereas early work focused on static network structures, there is increasing interest in identifying how networks change over time (Holme & Saramäki, 2012). In node-collected cases, edges must also be inferred over time (e.g. using sliding-window techniques), which involves a tradeoff between temporal resolution and estimate edge precision. Using more time-points to assist the edge inference will decrease the temporal resolution of the network, whereas using fewer time-points will entail a less precise estimate of the edge due to the instability of statistical estimates based on small numbers of samples. One must thus choose between increasing uncertainty or losing temporal resolution, both of which amplify uncertainty in the inferred edges within the temporal network. This will distort and blur subsequent properties derived from the representation, such as community detection.

Temporal community detection identifies fluctuating communities over time and can quantify changes in the interaction or groupings of nodes (Bazzi et al., 2016; Mucha, Richardson, Macon, Porter, & Onnela, 2010; Peixoto & Rosvall, 2017; Rosvall & Bergstrom, 2010). Community detection algorithms also contain uncertainty, as alternative methods will often produce slightly different results. Given that well-established static community detection algorithms can perform poorly when applied to complex real-world networks (Hric, Darst, & Fortunato, 2014), temporal extensions of these algorithms offer no inherent solution to uncertainty in the community detection step. Thus, two-step solutions to estimate communities from node-collected data (i.e. edge inference and then community detection) will propagate and smear the uncertainty that occurs at each step, warping both the communities and the interpretation of the dynamics. In order to quantify quick temporal properties in data, it is beneficial to understand the full effect of uncertainty.

The problems listed above can be mitigated using Temporal Communities through Trajectory Clustering (TCTC), which is designed to estimate temporal communities directly from node-collected data. TCTC bypasses the edge inference step

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and, instead, performs community detection in a single step directly from the time series. The solution here resembles ideas from the trajectory clustering literature (see Zheng, 2015 for a review), where their goal is to group trajectories in space and time. Importantly, TCTC is a general algorithm which can be of use to any dataset involving node-collected data where an underlying network structure is assumed.

There are numerous benefits to TCTC compared to existing methods. It can be used to discover the temporal and spatial properties of node-collected data. There are interpretable and concrete hyperparameters that can be meaningfully tuned to identify the optimal parameter settings for a contrast-of-interest. TCTC can also account for sudden spikes in noise that otherwise inherently bias methods that estimate network topology via a noisy edge-inference step. Finally, it identifies fluctuating temporal communities with a high temporal resolution, revealing new dynamic properties.

The article proceeds in the following manner: first, we introduce TCTC and demonstrate its utility through the recovery of expected time-averaged properties of task and resting state neuroimaging data. Thereafter we show that the temporal information within TCTC contains information that relates to single trial behaviour, revealing new information about ongoing temporal network configurations.

Results

Description of TCTC

We begin with outlining the conceptual innovation of TCTC. We will consider the input data of the algorithm to be a set of time series X consisting of N nodes and T time points. We further assume that the time series originating from recording activity of nodes in a network. The goal is to designate nodes into a community (or communities) at each time point. While current methods to estimate the time-varying communities from time series use a two-step process in which a network representation is first inferred through an edge inference step and followed by a separate community detection step, TCTC creates the communities from the time series in a single step (see *Fig. 1a* for an illustration how this approach relates to other approaches).

The TCTC algorithm identifies trajectories in a time series of nodes. If nodes are part of a trajectory, they get assigned to a community. For nodes to be part of the same trajectory, they must comply with four rules: distance, size, time and tolerance (*Fig. 1b*). The distance rule states that all nodes in a trajectory must be within ϵ of each other. The size rule states that all trajectories must contain σ number of nodes. The time rule states that all trajectories must be τ in temporal length. The tolerance rule states how many consecutive time-points the previous three rules can be violated while still allowing aa trajectory to persist. Together,

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

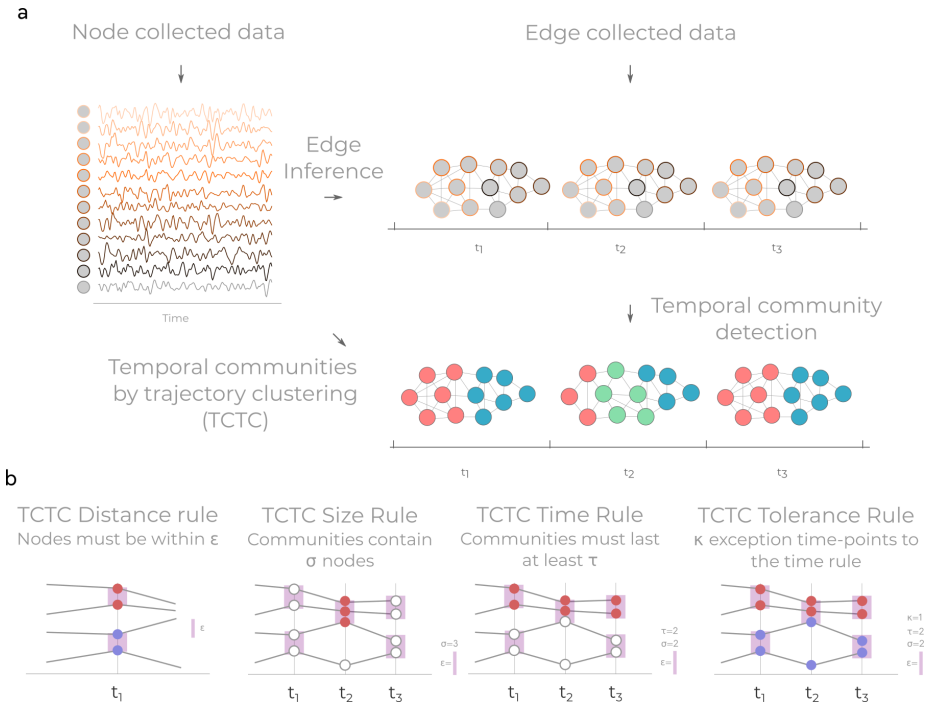

Figure 1: **An illustration of TCTC**. **a** How TCTC relates to other methods. Flow chart showing how temporal communities are derived contrasting TCTC from 2-step procedures on node collected data. Temporal communities can be derived on edge collected data in one step, node collected data in two steps via edge inference, and node collected data with TCTC in one step. **b** The four rules that define trajectories in TCTC. All four rules need to be obeyed in order to define a trajectory. Each rule shows four time-series with one or three discrete time-points. Purple rectangles show instances where the distance rule is successfully applied. White nodes indicate no trajectory is found.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

these hyperparameters define the minimum requirements for a community to exist within the data. These can be fit to find optimal hyperparameters by splitting the dataset into a training/test datasets (see N-back task below).

One of the advantages of TCTC is that each of the hyperparameters represent explicit properties of communities that are directly interpretable in and of themselves (Table 1). This helps to promote an understanding of the spatial-temporal properties of the communities, while also assisting in understanding optimization results (which are somewhat abstract in other approaches; e.g., the ω parameter in Mucha et al. (2010)). Finally, the four hyperparameters create the communities in a single step. Thus, all uncertainty about the communities is contained within these interpretable hyperparameters (See Supplementary Methods and *Fig. S1* for a walkthrough of each parameter and how if affects the community inference).

TCTC is a general solution that can be adopted by any discipline dealing with node-collected time series data (e.g., symptom networks in psychopathology, dynamic formation tracking in team sports, or environmental interactions). However, we will verify and demonstrate TCTC in relation to time series data found in neuroimaging. Our motivation here is that, in a recent review of community detection, the authors speculate that there will be an increase of algorithms designed upon the particular idiosyncratic properties of the data from that field (Fortunato & Hric, 2016). Here we will deal with several idiosyncratic properties of the data and collection design: node collected data, a low sampling-frequency-to-dynamics ratio (i.e. fMRI samples the brain quite slow compared to the dynamics of the brain), possibility to tune hyperparameters based on a contrast.

Parameter Description Unit ϵ The maximum distance between grouped time-series, given a clustering rule. Distance function used in this article: D1.
Distance between time-series (can be distance in amplitude or phase space). τ The minimum number of consecutive time-points the community must be present in data time points σ The minimum size of a community nodes per time point κ Number of consecutive "exception time points" that can exist in a row.
time points Table 1: Description of hyperparameters involved in TCTC

Exemplifying properties of TCTC

TCTC has additional features that make it particularly appealing in comparison to existing methods. First, when a node could be classified as being part of no trajectory with another node, TCTC classes it as not belonging in any community. This differs to other community detection approaches which often do not allow individual nodes to go unclassified. Here, an unconnected node is explicitly defined as having no apparent cooperation (given the parameters) at that time point. Another feature of TCTC is that it is a multi-label community detection algorithm (*Fig. S2*). This entails that a node can belong to multiple

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

communities at a single time-point and that communities can overlap. This is not common in many community detection algorithms, especially those currently applied in neuroimaging contexts. The advantage of a multi-label approach is that it does not force nodes to belong to a single community/process. Thus it allows, for example, a node to be connected to multiple processes that are processing unique information (i.e. different communities). Thus, a node that is accumulating information from multiple communities can become a member of each community with TCTC instead of forcing it to belong to a single community (or merging all communities into one).

To ease the interpretation of our new method, we define a set of additional metrics to quantify the time-varying communities TCTC. When validating TCTC in relation to time-averaged properties, we use the pairwise trajectory ratio (PTR) which is the percentage of time-points two nodes are in the same community. When illustrating the new properties that TCTC can discover, we relate the time-varying information in relation to a static community template. With this, static community co-occurrence (SCC) is the percentage of nodes from the static community combination that are in communities together (see Methods and *Fig. S2* methods for more information about SCC and PTR).

Validating TCTC on fMRI data

Here, wedemonstrate the validity of TCTC when applied to fMRI data. In order to validate the approach, we first consider whether the time-averaged communities from TCTC reveal static connectivity properties commonly found during rest with fMRI (Fox et al., 2005; Fransson, 2005). If this is the case, then TCTC is identifying properties that, while they may be fluctuating through time, when pooled together recreates the expected static relationship. We illustrate this both a resting-state fMRI and task fMRI datasets.

First, we use the resting-state data from the Midnight Scanning Club (MSC) dataset, which consists of 9 subjects (1 excluded due to poor signal quality) with 10 sessions each (Gordon et al., 2017). The TCTC parameters were set to: $\epsilon = 0.5$ (amplitude space), $\sigma = 5$, $\tau = 5$, $\kappa = 1$). These parameters were considered reasonable as they define communities that must be at least 5 time-points long, 5 nodes in size and all nodes are at least 0.5 standard deviation away from each other (see below for fitted hyperparameters).

We found a clear relationship between pairwise trajectory ratio (PTR) and the static functional connectivity (*Fig. 2a*). Further, we averaged the PTR for each static network combination (Yeo et al. (2011)), and find a large similarity between the average PTR and average functional connectivity between the static communities (*Fig. 2bc*). Here we see that TCTC is identifying properties similar to resting-state networks. However, to verify that TCTC is indeed finding session-specific properties, we compare the PTR with the functional connectivity from (i) the same subject/sessions (median ρ: 0.54), (ii) other sessions from the

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

same subject (median ρ: 0.42), and (iii) other subjects/sessions (median ρ: 0.29) (*Fig. 2d*). Here we see higher correlations when TCTC is matched with the session's corresponding functional connectivity and decreases as the expected relationship between the variables decrease. In sum, when averaging over time, TCTC recreates expected connectivity patterns at rest.

We assumed the hyperparameters in the verification above and it revealed a correlation between time-averaged TCTC and the functional connectivity. This presents a desired property towards which the TCTC hyperparameters can be optomized: minimizing the difference between TCTC's PTR and the static connectivity template. In doing so, we can see the effect of changing the different hyperparameters (see *Fig. S3*) and optimize them accordingly. Performing optimization steps (see task-fMRI example below), or using values based on previous optimizations, is advised for future discovery applications instead of assuming the values we initially chose.

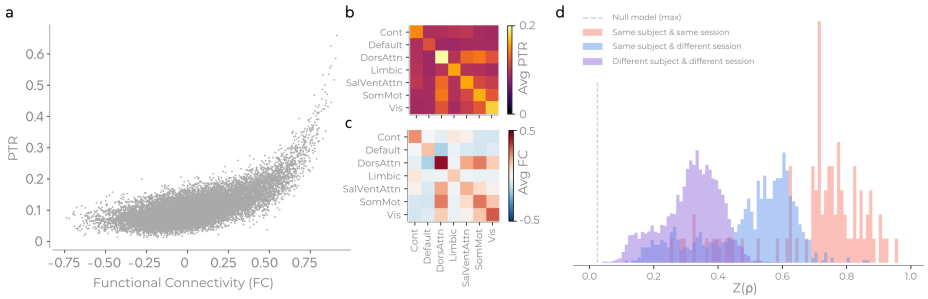

Figure 2: **Verifying TCTC on the MSC dataset**. **a** relationship of each edge's static functional connectivity and the ratio of time-points that two nodes get classed in the same community for an example subject/session; **b** Using a static brain network template, the average number of trajectories for each brain network combination from the example subject in *a*; **c** same as *b* but for functional connectivity. **d** Showing the distribution of the correlation (Fisher transformed Spearman values) in *a* for all subjects/sessions (same session); the distribution of pairwise trajectory ratio when correlated with the functional connectivity of different sessions from the same subject (same subject); the distribution of trajectory numbers when correlated with the functional connectivity of different subjects/sessions. All possible pairings for the three conditions were made in *d*. Dashed lines in *d* is the maximal value when permuting the functional connectivity edges that get paired with the PTR 100 times for each subject.

The foregoing analysis validates that TCTC is sensitive to time-averaged signal within its session compared to other sessions. Next we examined whether it is sensitive to expected time-average fMRI signals when performing a task. Here we use the data from the N-back task within the Human Connectome Project (HCP) data (Barch et al., 2013; Van Essen et al., 2012) and apply a similar logic to the previous analysis, in which we verify that TCTC identifies expected time-

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

averaged differences. The hyperparameters were optimized to find differences between 2-back and 0-back blocks in a training dataset (n = 50) and applied to a seperate test-dataset of (n = 50, see methods). The hyperparameters were: ϵ 0.52 (phase space), σ: 6, τ: 4, κ: 2. Using similar logic to the previous verification, we consider whether there are time-averaged differences in communities that resemble expected differences in an N-back task (e.g. Barch et al. (2013); Finc et al. (2017)). To quantify the difference between the conditions, the PTR differences between 2-back and 0-back on the test-dataset were used to identify which pairs of nodes that often ended up in the same community for a specific condition.

There were significant differences between the 0 and 2-back blocks on the test dataset (*Fig. 3ab*, NBS statistics, $p<0.001$, cluster threshold: 2). The communities identified in the 2-back block relate more to attentional, visual and control areas, whereas those identified in the 0-back task relate to the default mode network (*Fig. 3cde*). This is similar to expected findings between N-back fMRI tasks (Barch et al., 2013; Finc et al., 2017). Considering the subset of PTR combinations that were more frequent in the 2-back block, we observed a sustained period of activation throughout the block where there are more nodes throughout the entire time series in the 2-back condition (*Fig. 3e*). The reverse trend exists for the PTR during the 0-back block. This demonstrates that the result in *Fig. 3ab* are not driven by a handful of time-points, but instead are sustained throughout the blocks. However, this does not entail that there is no variance in the time-varying communities during the block, merely that the time-averaged differences are reflected throughout the block.

TCTC identifies fluctuating communities that have single trial properties

The previous section showed that that the average TCTC information contains relevant signal both in regards to task differences and subject properties during rest; however, the temporal specificity of TCTC also suggests that it may be useful to identify event-related effects with greater temporal precision. To this end, it is important to test what type of temporal information exists in the data. There are four possibilities regarding the type of fluctuations that the temporal information contains: (i) only noise, (ii) no temporal fluctuations, (iii) contain consistent block-specific temporal properties across subjects, or (iv) sensitivity to single-trial properties. We return to the HCP WM dataset 2-back blocks from the HCP dataset and first proceed by considering showing that there are both temporal fluctuations and these are not identical across subjects. Finally, we demonstrate that the temporal community fluctuations correlate with single trial behavioural properties, and thus showing that they are not merely noise.

When averaging over subjects, static community co-occurrence (SCC) 2-back values were found to not fluctuate across the entire time series (*Fig. 4a*), which may give the impression that there are indeed no interesting fluctuations in

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

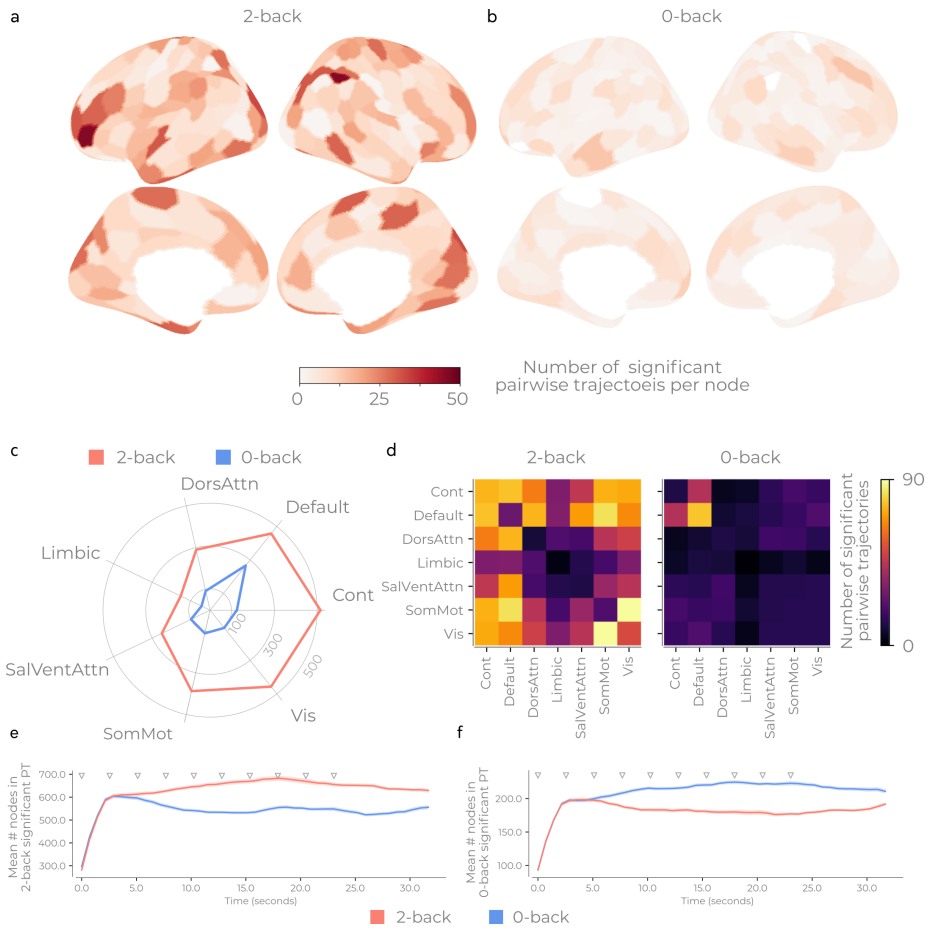

Figure 3: **Verifying TCTC on HCP N-back task**. **a** Differences between the number of trajectories assigned to a node differs between the 0 and 2 back tasks. **b** Placing the total trajectory differences between the two tasks onto the static information for the significant trajectories n the 2-back condition. **c** Same as *b* but tor the 0-back condition. **d** Displaying when in time the two sets of communities in **b** are present for both tasks. **e** Number of pairwise trajectories through time within the mask created from the trajectories that were significantly more present in 2-back. Triangles show stimulus onset times. **f** Same as *e* but for the mask created using the 0-back trajectories that were significantly more present.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TCTC. However, when looking at a specific block of a single subject (*Fig. 4b*, randomly chosen), we see that this subject's SCC values fluctuate over time. To show this is consistent across the dataset, we show that the average SCC standard deviation of each block is considerably larger and more varying than the standard deviation of the average SCC (*Fig. 4c*). If the SCC time series showed identical fluctuations across subjects/blocks, these would be similar. In sum, there are fluctuations in SCC and these fluctuations are not similar across subjects. Additionally, when considering resting state data from the MSC dataset and consider how the TCTC communities differ the distance between the static community template, we observe a heavy tailed distribution (*Fig. S4*) suggesting there are moments when the communities differ from the static template. Together, these results leave the possibility that the TCTC fluctuations are either noise or have single trial properties.

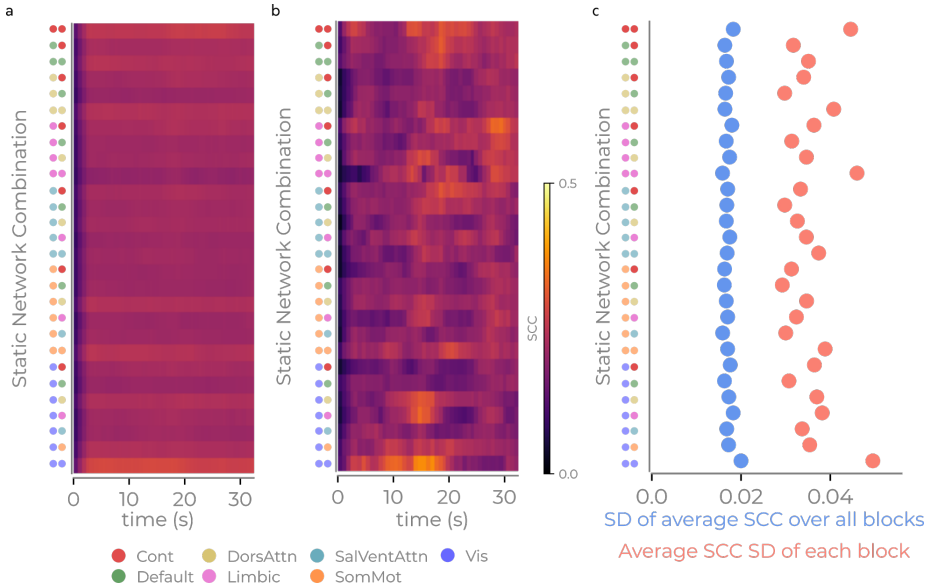

Figure 4: **Temporal properties of temporal communities of 2-back blocks. a** Average SCC (over all blocks and subjects) for each static network network combination. **b** SCC for a randomly chosen subject/block, illustrating the temporal properties of SCC. **c** The standard deviation of the average SCC (blue) and the average standard deviation of each SCC subject/block for each static network combination. This shows that each SCC time series of each block is more reflective of *b* than *a*.

After having established there are indeed temporal fluctuations in the TCTC data, we are left with the possibility that the temporal information contains either interesting signal or noise. In order to illustrate the properties of the temporal information, we derive SCC values relative to stimulus onset for each 2-back trial and correlate this with single trial response times and accuracy. We

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

first reduced the 28 network configurations to five PC components (*Fig. 5ab*, accounting for 74.5% variance). This is to collapse the number of features and mitigate the problem of multicollinearity in the subsequent statistical models. The PC components reflect different temporal community configurations which are summarized based on their static temporal profile. We then derive PC estimates per trial in relation to each stimulus onset for each second following stimulus onset (0-10 seconds) by using a weighted average of the sampled time-points. This is to ensure that we are not merely modelling the stimulus jitter. The earlier time-points will correspond to pre-stimulus activity due to BOLD sluggishness (we interrupt any result where t<4 to be prestimulus activity). Using a multi-level Bayesian model, allowing intercepts to vary for each subject/block, we estimated the posterior distributions for each of the PCs for each time-point following stimulus onset. Separate models were created for for each time-point. Three models were evaluated: reaction time (linear model), accuracy (logistic model, 1=correct, 0=incorrect) and response (logistic model, 1=response present, 0=no response). We will then highlight the peaks/troughs in the posterior distributions through time where the credible intervals are 90% above/below zero. Finally, will bring select the best features across the multiple time-points together into one model and verify this against an unseen dataset. All priors were set to be weakly informative (see Methods for further details). Posterior distributions were estimated using 30,000 MCMC draws in total from 3 chains after each having 1,000 tuning samples.

The five PC components each represent different temporal network configurations expressed along the static template dimensions (*Fig. 5ab*). PC1 shows a general increase in all communities. PC2 is a community configuration containing more nodes from the visual network (both with other visual network nodes and with attention and sensorimotor networks). PC3 is marked by a reduction in visual network communities (both within the visual network and between attention and control networks). There is also an increase in default mode communities (both within and between other networks). PC4 shows nodes in the cognitive control network having an increase their number of communities (both within and between). PC5 shows an increase in communities between cognitive control, dorsal attention and limbic networks. The visual network also decreases in communities, except with the dorsal attention. Together, these five components show a diverse number of community assignments.

The results reveal that the temporal network configurations are associated with behavior differently depending on when they occur (*Fig. 5cde*). An increase in PC3, with reduced connectivity in visual and control networks, prior to stimulus onset is associated with quicker reaction times (*Fig. 5c*). However if this network configuration persists later in the trial, then PC3 is associated with missed responses (*Fig. 5e*). Multiple behavioural roles are also seen for PC2 and PC4. When PC2 is engaged early in the trial, subjects were more accurate (*Fig. 5d*) whereas the same component was associated with longer reaction times later in the trial (*Fig. 5c*). PC4 correlates with quicker reaction times both at an early and later time-point (*Fig. 5c*). PC4 also correlates more accurate response but

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

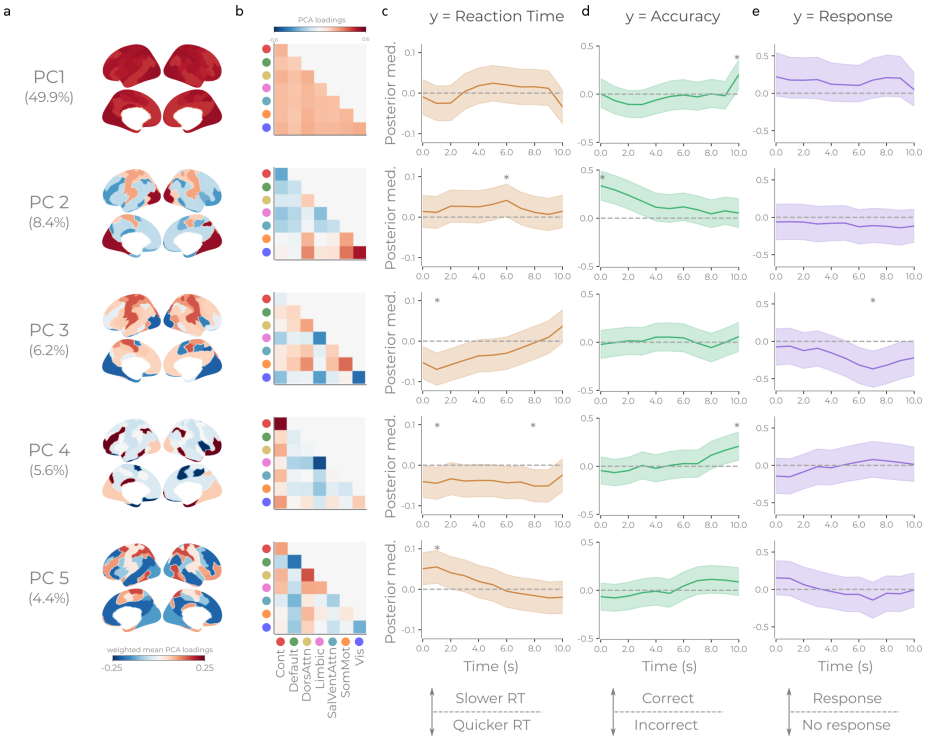

Figure 5: **Snapshots of temporal community configurations correlate with single-trial behaviour.** **a** average is a weighted average where the "within" static network loading is weighted as much as all the "between" static networks together. **b** PCA loadings when reducing the network combination for each trial to 5 PCA components for the SCC values at around each trial onset. **c** Linear model where single trial reaction times are predicted with the SCC PC components. Lines show posterior medians of each PC for Bayesian models fitted for each time-point relative to stimulus onset. Shaded region shows 90% credible interval. Asterisks depict max/min median posterior value through time where the entire 90% credible interval peaks either above/below zero. **d** Same as *c* but for a logistic model modelling accuracy. **e** same as *c* but a logistic model modelling whether there was a response or not.)

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

later in the trial (*Fig. 5d*). Additionally, we see that an increase in PC5 early in the trial leads to longer reaction times (*Fig. 5c*). Here we see an example where two network configurations correlate have opposite effect on behaviour at a specific time (see PC3 at t=1). Finally, we also see that two network configurations correlate with behaviour concurrently. This we see post-stimulus accuracy with PC1 and PC4 (*Fig. 5d*).

A detailed picture emerges in which different network interactions, at various times, can explain multiple behavioural measures. This confirms that TCTC is sensitive to behaviourally-relevant temporal information within neuroimaging data. However, as each time-point has received its own model, we have yet to show if multiple time-points together explain single-trial behaviour and, at present, could increase the number of false positives. We select the features where a network combination peaks where 90% of the credible interval is above/below 0 (*Fig. 5cde*) into one model per behaviour. We calculate leave-one-out (LOO) model evaluation for each feature independently and a combined model. The LOO reveals that combined models were the best performing (see Table S1-S3).The selected models for verification (see below) for the three behaviours are the best performing models. For reaction time and accuracy, this is the combined model of features across time points. For response, this is was PC3 only at t7 shown in Figure 5. Posterior distributions for the combined models are shown in *Fig. 6abc*. In sum, the different TCTC configuration at multiple time-points together can explain single trial behavioural properties.

Finally, in order to verify the selected models for each behaviour, we sample from the posterior distribution and compare the sampled data to both the original data (posterior predictive checks) and to an unseen verification dataset (prediction). For the linear model for reaction time, the mean and interquartile range (IQR) of the simulated data were compared with the original data (mean: p=0.49, IQR: p=0.64) indicating a good model fit which should generalize to new data. With the separate verification data, there was a correlation between the average posterior sampled data and the new verification data's reaction times (median of posterior=0.18, 90% CI=[0.14, 0.23], 100% posterior above 0, *Fig. S3*). This shows that the model generalizes to new data but also emphasises that the model is only capturing part of the variance of single trial reaction times. In regards to the logistic models, we calculated the weighted F1 score of classifying a trial to an outcome. Due to the small number of errors and missed responses, we first identify an appropriate cutoff threshold for prediction by inspecting the ROC curve after sampling the posterior distribution. We then use this cut-off on when sampling from the posterior and comparing with the verification dataset. Both accuracy and response models had a high weighted F1 score (accuracy: original data: 0.80; verification data: 0.70; response: original data: 0.89; verification data: 0.83). In sum, all models show their posterior distributions can model new unseen data.

In sum, we have demonstrated with TCTC that multiple network configurations are behaviourally relevant at specific moments and other configurations can be

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

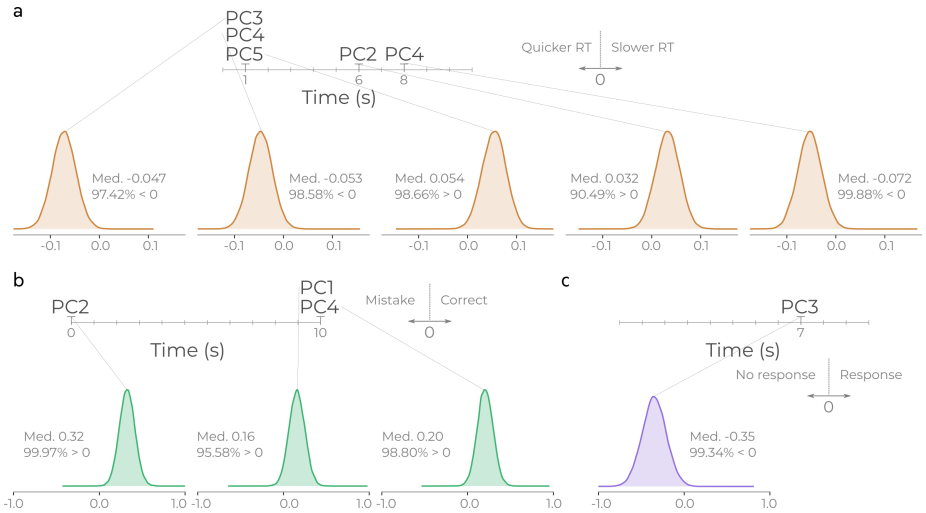

Figure 6: **Posterior distributions of multi-time point models shows multiple temporal network configurations correlate with single-trial behaviour. a** Shows posterior distributions of 5 PC component selected for reaction time model. Lines from each distribution point to which PC component and when in time. **b** same as *a* but for three selected components for accuracy. **c** same as *a* but for the one selected component for response model

equally or more important at other moments. This signifies that the ongoing community configurations are in flux and these configurations are important for efficient transferral of information across the brain. This demonstrates a continual interplay of interactions across the traditionally static brain networks when performing a task.

Discussion

Here we have introduced TCTC, a multi-label community detection algorithm designed for node collected data, and demonstrated the utility of the algorithm on two separate functional neuroimaging datasets. The TCTC represents a substantial improvement over current state-of-the-art techniques for community detection on temporal node-collected data. Critically, it is able to estimate community structure directly from time-series, without requiring additional estimates of network edges. To validate this approach, we have shown: (i) that the hyperparameters from TCTC are flexible and directly interpretable, (ii) that expected time-averaged results can be recovered from time-averaged communities, and (iii) that the temporal information identified in the communities correlates with single trial behaviour on a challenging cognitive task. Thus, the TCTC represents a robust algorithm for probing how and to what degree the interacting

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sets of communities cooperate through time in relation to a task.

One of the most useful features of TCTC is its ability to detect differing temporal and spatial scales of the communities when there is sufficient data to make a training/test split. From there, TCTC can be used to determine the temporal-spatial scale of the data relative to any type of contrast (tasks, behaviour, genetic differences, etc.). In theory, any community detection algorithm can be optimized on a training dataset to optimize the parameters in a similar way. However such optimization procedures are not typically performed, especially in neuroimaging experiments. Instead, they often select a value (usually a default value) and show that results are relatively consistent after slightly varying the parameters (e.g. Betzel, Fukushima, He, Zuo, & Sporns, 2016). This preference for heuristics over optimization may arise because the parameters in other approaches are often abstract; e.g., finding a Louvain γ of 0.7 over 0.8 is less concrete than finding that τ parameter in TCTC should be 4 instead of 3 — the latter has a much clearer interpretation. Thus, optimizing the hyperparameters is both useful and understandable. Indeed, the process of optimizing TCTC parameters, because they are directly interpretable, can be informative.

We have also demonstrated novel findings with TCTC regarding temporal communities and neuroimaging. Previously, researchers have quantified how average performance relates aggregate measures of temporal communities (e.g. Bassett et al. (2011); Shine, Koyejo, & Poldrack (2016); Saggar et al. (2018)) comparing rest to task (Mattar, Cole, Thompson-Schill, & Bassett (2015)), all using two-step methods. Usually behavioural correlates are based on average behaviours (e.g. Saggar et al. (2018)). To our knowledge only Fransson and colleagues (Fransson, Schiffler, & Thompson, 2018) have found correlations between single trial reaction time/accuracy data and temporal network properties, but some significant differences exist with their perspective, including assuming both static communities and that a temporal network time-series has a single behavioral property. Here we have demonstrated that roles can change during a time series and the importance of temporal communities.

We have demonstrated that temporal communities, at different time-points, account for single trial variance in behaviour. This is a contrasting perspective of brain processing from many common practices today within neuroimaging where the aim is to identify brain regions, patterns, networks, or network configuration that are considered "on" or "more on" during a task condition. Instead, it suggests we should view a cognitive processes in terms of information flow in the brain occurring between communities that merge and split based on the task at hand. If the right nodes interact and at the right time, the correct information is able to flow around the brain and, in turn, will lead to greater accuracy and quicker reaction times. To perform a task, the dynamic coordination of multiple brain regions can affect performance but only at the correct time, otherwise they can be detrimental (e.g. PC3 must occur at the optimal time to have a quicker reaction time and not miss the response). Such temporal zones of useful connectivity configurations leads to many new possible hypotheses regarding

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

how large scale networks should be quantified and how they get attributed for different cognitive processes. Primarily, this perspective suggests pivoting the field away from identifying brain areas/networks/network configurations that are merely "on" or "off" more during a task, and towards identifying the temporal-spatial configurations of networks as they facilitate information flow in the brain. In sum, the results from TCTC opens up an avenue of research questions to explore quick changes in time-varying communities in neuroimaging contexts.

This temporal-spatial configuration perspective can be integrated with the low dimensional manifold perspective (Shine et al., 2019) whereby slow shifts across tasks occur in the underlying network architecture and some of the PC's found here show similarities to those in Shine et al. Here however, we see both quick fluctuations in community configurations which accounts for single trial behavioral variance and different also different configurations occurring different block types. TCTC fluctuations are fast enough that they could be occurring along the ongoing low dimensional manifolds shifts observed by Shine and colleagues to occur between tasks.

Multiple modifications could be made to TCTC, which may be appropriate for different use-cases. At present, lagged relationships cannot be detected, but this could be rectified by using dynamic time warping as a distance measure instead of D_1 used here. Other preprocessing steps exist the trajectory clustering literature, such as an initial compression of the time series, which can speed up calculations. There are also multiple additional clustering algorithms from the trajectory clustering literature (convoys, swarms) which could be applied (Zheng, 2015). Another possible extension is to add the time-series of confounds into the community detection algorithm. If a community also contains these confounds, this community should be rejected.

One final noteworthy property of TCTC, although was unexplored in this article, is that the multi-label communities can be converted to create binary time-varying connectivity matrices with no loss of information. This opens up additional possibilities for time-varying connectivity through trajectory clustering (TVCTC) and perform analyses beyond community detection.

Community detection is an important part of network theory. In complex networks like the brain, there is ample evidence to conclude that functional network structure fluctuates over time. Here we have presented a way for these communities to be derived. Further, TCTC is not merely a "black box" method which produces a community vector, but rather the hyperparameters are concrete in their formulation to help shape what type of communities are identified. Indeed the hyperparameters, if applied to multiple datasets, could reveal interesting properties about the spatial and temporal scales of different task contexts.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Description of TCTC

Let X be a discrete time series consisting of N nodes and T time points. X can express either the amplitude or instantaneous phase of the nodes. The goal is to create communities for each time point. In TCTC, communities are identified between groups of nodes if the nodes are part of a trajectory. To identify trajectories, there are four rules each with their own hyperparameter: a distance rule, a duration rule, a size rule, and a tolerance rule.

We use the notation C_t^A to notate "Community A at time-point t" which consists of a set containing node indices that signify the nodes belong to that community.

Distance rule

The distance rule specifies how far the time series in X can be from each other in order to be considered part of the same trajectory. Given some distance function, $D(X)$, the hyperparameter ϵ specifies the maximal allowed distance that time points can be from each other in order to be part of the same trajectory. When ϵ is small, only time series with very similar values with get grouped together. As ϵ increases, more dissimilar time series will get clustered together. This rule can be formulated as:

$$\forall_{i \in C_t^A} \forall_{j \in C_t^A} D(X_{it}, X_{jt}) \leq \epsilon$$

Where $\forall i$ indicates "for all i". This entails that the maximal distance between any node in a trajectory is ϵ. This formulation creates an analytic uncertainty of all the nodes within a community. For the distance function, we use D_1 distance whenever X consists of amplitudes and, when X contains phase information the distance function is: $D = |X_{jt} - X_{jt}| \bmod 2\pi$ (i.e. remainder of D_1 after dividing by 2 π). The above formulation is similar to "flock clustering" from the trajectory clustering literature (Zheng, 2015).

All the uncertainty of community derivation is defined by the hyperparameters, the upper-bound of the uncertainty is easy to calculate. For example, all communities must contain nodes that are ϵ (see below) distance away from each other (except tolerance time-points). This allows for a maximum bound of uncertainty to be quantified as it is intertwined with the hyperparameters. This is not possible with the two-step processes that have unrelated uncertainty in both steps.

Duration rule

The duration parameter (τ) specifies the minimum length of the trajectory. This entails that the nodes that are a member of C_t^A are also a member of C_{t+1}^A

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and all subsequent time-point up until $C^A_{t+\tau-1}$. That is to say the community obeying the duration rule must follow:

$$\forall_{i \in C^A_{t_0}}, i \in C^A_{t_0+1}, ..., i \in C^A_{t_0+\tau-1}$$

Where t_0 signifies the first time-point in a trajectory.

Trajectory size

The size parameter (σ) specifies the minimum number of nodes that are part of the trajectory. This entails that there must be at least σ nodes belong in C^t_A. Explicitly, a community must follow the following size rule:

$$|C^A_t| \geq \sigma$$

where $||$ indicates the number of elements within the set.

Tolerance rule

The tolerance rule specifies how many consecutive exceptions are allowed where the distance rule or size rule fails. The idea here is that, if a brief spike in noise effects one or more of the time series, this will interrupt the trajectory. We can amend the duration rule and add a parameter κ which allows for the number of consecutive exceptions:

$$\forall_{i \in C^A_{t_0}}, \exists_k (i \in C^A_{t_0+1+k}), ..., i \in C^A_{t_0+\tau-1+N_\kappa}, \text{ where } k = 0, 1, ...\kappa$$

Where N_κ is total the number of time-points that were tolerated. If $\tau = 3$ and $\kappa = 1$ then it is possible for there to be two instances where the tolerance rule can be applied (at $t_0 + 1$ and $t_0 + 3$), then $N_\kappa = 2$. This results in all members of C^A being present at t_0, $t_0 + 2$, $t_0 + 4$.

fMRI data

Midnight scanning club data

10 subjects with 10 resting-state sessions (818 time-points) from the Midnight Scanning Club (MSC) dataset (Gordon et al., 2017). One subject (MSC08) was deleted as is known to be noisy. The preprocessed data as outlined in (Gordon et al. (2017)) and is available on OpenNeuro was used. The only exception was that 200 parcels were created from the Schaefer atlas (Schaefer et al., 2018) and the Yeo 7-community static network parcellation (Yeo et al., 2011). Static functional connectivity was calculated with a pearson correlation across each

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pairwise combination. For TCTC, the time series were first standardized to have a mean of 0 and standard deviation of 1.

N-back task

100 subjects from the Human Connectome Project N-back task while recording fMRI (100 unrelated subject release, TR=0.72, minimal preprocessed data used) (Glasser et al., 2013; Van Essen et al., 2012). The LR encoding dataset was used throughout the paper. The RL encoding was used in verifying the Bayesian models. The same 200 ROIs and 7 static network parcellation as the MSC dataset were derived from the greyordinatesx(Schaefer et al., 2018). After the HCP minimal preprocessing and regressed out 12 movement parameters, framewise displacement and global ROI mean. Additional preprocessing steps were performed on this data, as no temporal filtering has been applied on this dataset. Scrubbing was used to remove any time points that had a framewise displacement (FWD) value greater than 0.5. Missing data was simulated with a cubic spline to create continuous time series. If more than 20% of a subjects data was simulated, than the subject would be removed. No subjects were removed. The data was band-passed between 0.01 and 0.12 Hz and the data was converted to instantaneous phase. Each subject performed 8 blocks (four 0-back and four 2-back). Additional preprocessing was done using nilearn (Abraham et al., 2014) and teneto (V0.3.4) (Thompson, 2019).

The LR encoded data was split into a training and test datasets, each containing 50 subjects. Each subject had four 2-back and four 0-back blocks where stimuli were presented every 2.5 seconds.

Furthermore, in order for communities that start at the final time-points to have enough time to terminate, the end of blocks were padded with τ time-points.

TCTC parameters

For the MSC resting state data, preset parameters were chosen. These were: τ: 5, ϵ: 0.5 (amplitude space), σ: 5, κ: 1. One of the reasons for choosing preset here is to demonstrate how these parameters are interpretable and verify them on the static properties without optimizing to the static properties. Here we have stated that to be a community they must: last for 5 time-points, consist of at least 5 nodes, and those 5 nodes must always be within half a standard deviation of each other. Finally there may be violations of one time point to the above rules. How these parameters shape the community detection algorithm is straight forward given the definitions. However, these parameters are not optimized. A possible future extension is to add an unsupervised optimization of these parameters or optimize to the static properties (see Results).

For the N-back task data, the objective function was defined to maximise the hamming distance between binary trajectory clustering matrices (i.e. if two nodes are in a trajectory together at t then they receive a connection of 1) during the

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

0-back vs 2-back blocks for each time point. Each block was 32.4+τ seconds long. The optimization function was run on 25.2 seconds long (35 time-points) starting after 7.2 seconds after the block began (10 time-points). The offset was to avoid training on any spillover from previous block due to the sluggishness of BOLD signal in fMRI.

The hyperparameters were chosen by optimizing on 50 subjects in a training dataset. 50 iterations of Bayesian optimization implemented in *scikit-optimize* (V0.5.2) (Head et al., 2018) were used to search for appropriate hyperparameters. The parameter search space was: σ: 3 to 20 (node size), ϵ: 0.01 to $\pi/6$ of phase space, τ: 3-15 (time). κ was set to be between 0 and 2. All hyperparameters were sampled uniformly except *epsilon* which was sampled from a log-normal distribution. All results are shown on the application to the test dataset.

The goal of the optimization was to minimize the following equation:

$$O = \frac{2}{N_{2B}^2 - N_{2B}} \sum_{i=1,j=i+1}^{N_{2B}} D_H(U_i^{2B}, U_j^{2B}) + \frac{2}{N_{0B}^2 - N_{0B}} \sum_{i=1,j=i+1}^{N_{0B}} D_H(U_i^{0B}, U_j^{0B})$$

$$- \frac{1}{N_{2B} N_{0B}} \sum_{i=1}^{N_{2B}} \sum_{j=i}^{N_{0B}} D_H(U_i^{0B}, U_j^{0B})$$

Where U is the upper-triangular of each temporal snapshot of the binary trajectory matrix (dimensions: node, node, time) where 1 signifies a trajectory is present. D_H is the hamming distance. 0B and 2B indicate 0-back or 2-black conditions and i and j are the condition index of U. N_{2B} is the number of blocks of the 2-back condition were performed (here four). To minimize O, this entails that average difference between "within 2-back" and "within 0-back" blocks should have a small hamming distance and the average difference "between 2-back and 0-back" blocks should be as large as possible. On each optimization iteration the average O for all subjects in the discovery dataset was calculated. A termination rule was also implemented when there were 10 blocks that failed to find any trajectories, that parameter combination was considered "bad" and ended with a value of an objective function value to 1 (this was done for computational speed).

After optimization the parameters derived on the training dataset were: τ: 4, ϵ 0.52, σ: 6, κ: 2.

Note, on the MSC dataset we found communities based on the amplitude of the nodes. For the N-back we identified communities based on the phase of the time series. We have done this to illustrate the two different possibilities for TCTC. Our preference is to use phase space, especially for task data, as any mean non-stationary occurring in the data will affect the amplitude space communities more than phase space.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Community detection metrics

Pairwise trajectory ratio (PTR)

To summarize the amount of interaction between the identified communities through TCTC and the static network template, we derived the *pairwise trajectory ratio*. For each node pairing, we count the percentage of time-points that two nodes appeared in at least one community together.

Static community co-occurrence (SCC)

To summarize the amount of interaction between the identified communities through TCTC and the static network template, we derived the *static template co-occurrence*. For each static community pairing, we count the percentage of nodes in all TCTC communities that intersect with the static functional network template from all possible nodes. Namely:

$$\mathrm{SCC}_{n,m,t} = \sum_{A,t} \frac{|C_t^A \cap (S_n \cup S_m)|}{|S_n \cup S_m|}$$

Where S_n and S_m are sets containing nodes indices for the static community partition for static network indices n and m. This measure is thus reflective of the overall cooperation, between the two static communities. When the term "within-SCC" it refers to the $\mathrm{SCC}_{n,n}$ and the "between-SCC" is the average $\mathrm{SCC}_{n,m,t}$ over all m when $n \neq m$. Note that this measure entails that the multiple time-varying communities can be comprised of nodes originating from the same static network communities.

Statistics

To illustrate that there was a difference between the two N-back blocks in the test dataset, the network based statistic (NBS) was used (Zalesky, Fornito, & Bullmore, 2010). The pairwise trajectory ratio was averaged over time. The aim of NBS is to find clusters of edges that significantly differ between conditions. Each block (4 per condition, per subject) were permuted, shuffling the condition-membership to create the permuted distribution. The trajectories from each block were averaged over time entailing that any difference found is a time-averaged difference between the conditions. 1,000 permutations were performed. The cluster threshold was set to 2 and significance was considered if $p<0.001$. This defines a set of edges where the frequency of trajectories over time between different nodes were significantly different between the communities.

We used a hierarchical Bayesian model to quantify the difference between single trial behaviour (reaction times, accuracy, and response). SCC's, these were whitened, and 5 PCA components were derived expressing temporal network

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

community configurations at each time-point during the block. The sampling of fMRI volumes does not correspond to stimulus onset. To account for this and to make sure we are not merely fitting the stimulus jitter, a weighted average of the two encompassing PC time-points was used to align all trials to the same temporal offset. Each statistical model was run for the stimulus onset-locked PCs and up to 10 seconds afterwards. As different individuals have different reaction times for different blocks (as each block had different stimuli types), each block got its own intercept modelled.

The statistical model that models single trial reaction times of correct trials from the community snapshots was specified as:

$$y_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta_1 \text{PC1}_{t,i} + \beta_2 \text{PC2}_{t,i} + \beta_3 \text{PC3}_{t,i} + \beta_4 \text{PC4}_{t,i} + \beta_5 \text{PC5}_{t,i}$$
$$\alpha \sim \text{Normal}(\mu_{\text{block}}, \sigma_{\text{block}})$$
$$\beta_{1-5} \sim \text{Normal}(0, 1)$$
$$\sigma \sim \text{HalfCauchy}(5)$$
$$\mu_{\text{block}} \sim \text{Normal}(0, 1)$$
$$\sigma_{\text{block}} \sim \text{HalfCauchy}(1)$$

Where y_i was the reaction times. A Box-Cox transform was applied to the reaction times in order to transform them towards a Gaussian distribution (λ = -0.011). All priors are weakly informative priors. The reaction times and PC components were standardized so the β values are on a comparable scales. This hierarchical models an intercept (α) for each of the blocks. MCMC was performed using pymc3 (Salvatier, Wiecki, & Fonnesbeck, 2016). 10,000 samples were drawn from 3 separate chains (1,000 tuning samples for each) using the no-u-turns sampler (NUTS).The model was run for 11 different values of t (t=0, to t=10). We interpret t<4 to reflect pre-stimulus activity due to the sluggishness of the BOLD response.

For modelling both accuracy and response, a similar model was applied as above, except with small modifications to make the model logistic instead of linear to account for y being binary values where 0 was an incorrect or miss trial and 1 was a correct trial:

$$y_i \sim \text{Binomial}(\text{invlogit}(\mu_i))$$
$$\mu_i = \alpha + \beta_1 \text{PC1}_{t,i} + \beta_2 \text{PC2}_{t,i} + \beta_3 \text{PC3}_{t,i} + \beta_4 \text{PC4}_{t,i} + \beta_5 \text{PC5}_{t,i}$$
$$\alpha \sim \text{Normal}(\mu_{\text{block}}, \sigma_{\text{block}})$$
$$\beta_{1-5} \sim \text{Normal}(0, 1)$$
$$\mu_{\text{block}} \sim \text{Normal}(0, 1)$$

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\sigma_{\text{block}} \sim \text{HalfCauchy}(1)$$

Evaluation checks of MCMC were done through manual inspection and checking that the Gelman-Rubin statistic was close to 1. For model selection the leave one out information (LOO) criteria was used. The best LOO model for each behaviour was then verified with a unseen dataset.

Verification of models involved sampling from the posterior distribution and comparing to the verification dataset. The verification dataset was RL encoding of the WM HCP dataset using the same subjects. All transforms (PCA, Box Cox, standardization) were based on the original data. 10,000 samples for each value were drawn from the posterior. For the linear model (reaction time), posterior predictive p-values were calculated between the simulated value and the original data for the mean and IQR. With the verification dataset, a linear model was fitted between the simulated data and unseen reaction times sampling all priors and distributions for similar to the linear model outlined above (where the model was specified as: VerificationRTs $\sim \alpha + \beta$ PosteriorRTs, with all priors being weakly informative and the independent and dependent variables were standardized). For logistic models a predictive threshold was calculated after viewing the ROC curves and selecting the largest point that corresponded to: (1-false positives)+(true positives) on the original data. This threshold was then applied to the verification dataset and the weighted F1 score was used to evaluate the predictive accuracy of the models.

Data availability

Code for TCTC is implemented in Teneto (https://github.com/wiheto/teneto) from teneto v0.4.4 and onward. Docker container and scripts for the analyses used can be found at (https://github.com/wiheto/project_code/tctc_paper/). Data for N-back task can be found on the Human Connectome Project homepage (https://humanconnectome.org/) and MSC dataset is available on Open Neuro (https://openneuro.org).

Reference

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., ... Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*, 14.

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., ... Van Essen, D. C. (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage*, *80*, 169–189.

Bassett, D. S., Wymbs, N. F., Porter, M. a, Mucha, P. J., Carlson, J. M., & Grafton, S. T. (2011). Dynamic reconfiguration of human brain networks during

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

learning. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 7641–6.

Bazzi, M., Porter, M. a., Williams, S., McDonald, M., Fenn, D. J., & Howison, S. D. (2016). Community detection in temporal multilayer networks, and its application to correlation networks. *Multiscale Modeling & Simulation*, *14*, 1–41.

Betzel, R. F., Fukushima, M., He, Y., Zuo, X. N., & Sporns, O. (2016). Dynamic fluctuations coincide with periods of high and low modularity in resting-state functional brain networks. *NeuroImage*, *127*, 287–297.

Finc, K., Bonna, K., Lewandowska, M., Wolak, T., Nikadon, J., Dreszer, J., ... Kühn, S. (2017). Transition of the functional brain network related to increasing cognitive demands. *Human Brain Mapping*, *38*, 3659–3674.

Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, *659*, 1–44.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 9673–8.

Fransson, P. (2005). Spontaneous low-frequency BOLD signal fluctuations: An fMRI investigation of the resting-state default mode of brain function hypothesis. *Human Brain Mapping*, *26*, 15–29.

Fransson, P., Schiffler, B. C., & Thompson, W. H. (2018). Brain network segregation and integration during an epoch-related working memory fMRI experiment . *Neuroimage*, *178*, 147–161.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., ... Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–24.

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Petersen, S. E., Nelson, S. M., Dosenbach, N. U. F., ... Greene, D. J. (2017). Precision Functional Mapping of Individual Human NeuroResource Precision Functional Mapping of Individual Human Brains. *Neuron*, 1–17.

Head, T., MechCoder, Louppe, G., Shcherbatyi, I., Fcharras, Vinícius, Z., ... Fabisch, A. (2018, March). Scikit-optimize/scikit-optimize: V0.5.2. doi:10.5281/zenodo.1207017

Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics Reports*, *519*, 97–125.

Hric, D., Darst, R. K., & Fortunato, S. (2014). Community detection in networks: Structural communities versus ground truth. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *90*, 1–19.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mattar, M. G., Cole, M. W., Thompson-Schill, S. L., & Bassett, D. S. (2015). A Functional Cartography of Cognitive Systems. *PLOS Computational Biology*, *11*, e1004533.

Mucha, P. J., Richardson, T., Macon, K., Porter, M. a, & Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science (New York, N.Y.)*, *328*, 876–8.

Newman, M. (2010). *Networks. An introduction* (p. 772).

Peixoto, T. P., & Rosvall, M. (2017). Modelling sequences and temporal networks with dynamic community structures. *Nature Communications*, *8*, 1–12.

Rosvall, M., & Bergstrom, C. T. (2010). Mapping change in large networks. *PloS One*, *5*, e8694.

Saggar, M., Sporns, O., Gonzalez-castillo, J., Bandettini, P. A., Carlsson, G., Glover, G., & Reiss, A. L. (2018). Towards a new approach to reveal dynamical organization of the brain using topological data analysis Manish. *Nature Communications*, *9*, 1–14.

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, *2*, e55.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A., ... Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex From Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, *28*, 3095–3114.

Shine, J. M., Breakspear, M., Bell, P. T., Martens, K. A. E., Shine, R., Koyejo, O., ... Poldrack, R. A. (2019). Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nature Neuroscience*, *22*, 289.

Shine, J. M., Koyejo, O., & Poldrack, R. A. (2016). Temporal metastates are associated with differential patterns of time-resolved connectivity, network topology, and attention. *Proceedings of the National Academy of Sciences*, *113*, 9888–9891.

Thompson, W. H. (2019). Teneto. doi:10.5281/zenodo.2535994

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., ... Yacoub, E. (2012). The Human Connectome Project: a data acquisition perspective. *NeuroImage*, *62*, 2222–31.

Yeo, B., Krienen, F., Sepulcre, J., Sabuncu, M., Lashkar, D., Hollinshead, M., ... Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*, 1125–1165.

Zalesky, A., Fornito, A., & Bullmore, E. T. (2010). Network-based statistic: identifying differences in brain networks. *NeuroImage*, *53*, 1197–207.

Zheng, Y. (2015). Trajectory Data Mining: An Overview. *ACM Trans. On Intelligent Systems and Technology*, *6*, 1–41.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgments

WHT acknowledges support from the Knut och Alice Wallenbergs Stiftelse (SE) (grant no. 2016.0473, http://kaw.wallenberg.org). This Midnight Scanning Club dataset was obtained from the OpenNeuro database. Its accession number is ds000235. Human connectome data was provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University."

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figures

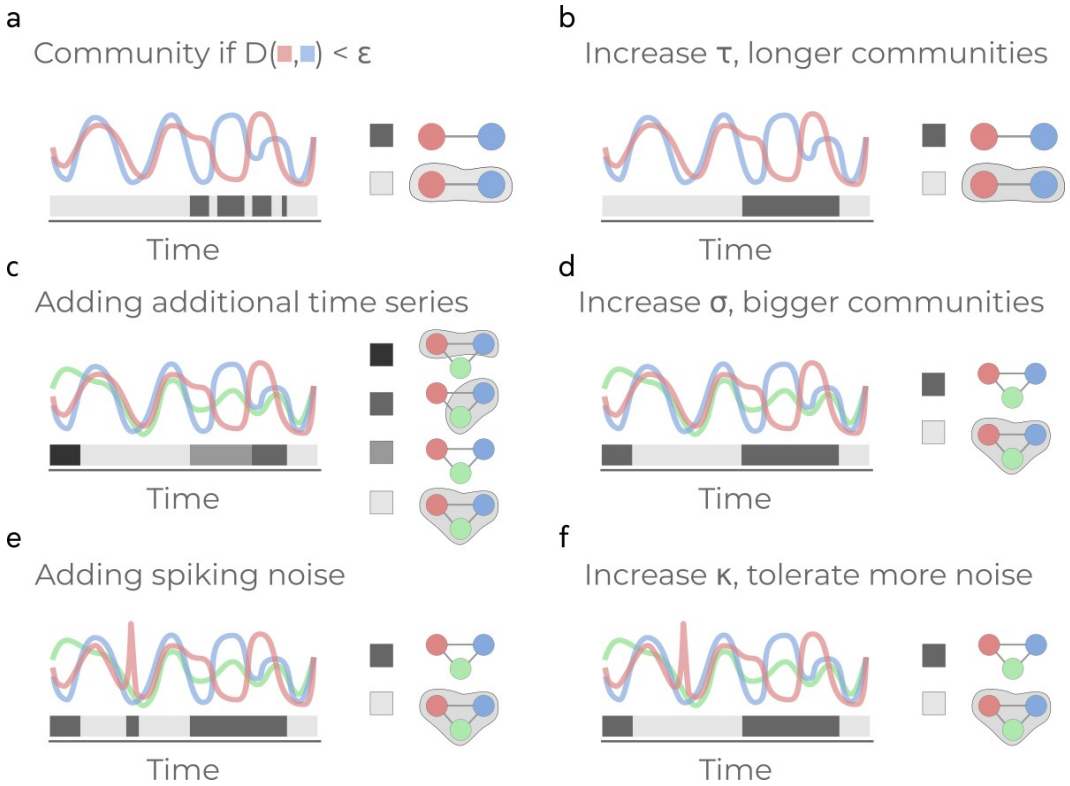

Fig. S1. Illustration of the different hyperparameters in TCTC and community inference from trajectories. Each panel shows time series which are assumed to be node collected data from a network. The coloured bar along the bottom indicates the community assignment for that time point. The right side of each panel indicates the community structure where grouped nodes are joined together with shaded background. **a.** Two time series where τ is low are considered members of the same community if the distance between the two time series is less than ε. **b.** What happens to *a* when τ is increased. **c.** Adding an additional time series to *b* and σ = 2. **d.** The effect on *c* when increasing σ. **e.** *c* has now been modified with a brief burst of noise. **f.** The effect to the communities in *e* when increasing κ. This step ignores κ number of time-points when the previous rules are not met.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

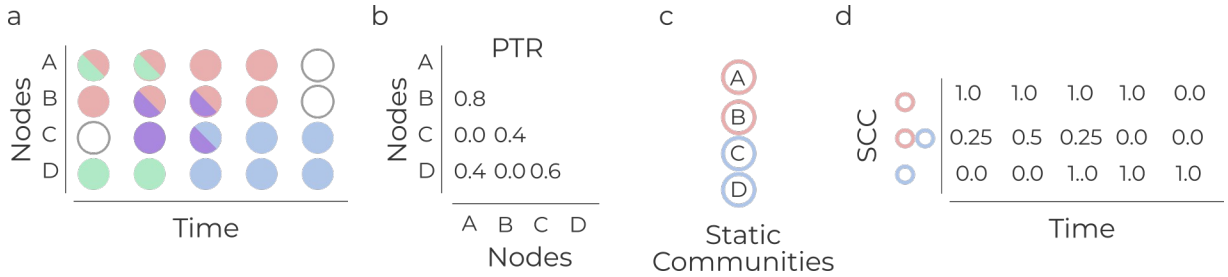

Fig. S2. Example of PTR and SCC to quantify the multi-label time-varying community assignments. **a.** an example of time-varying community assignment for five time-points. Time-points where a node has a white circle are when no community is assigned. Half circles indicate multiple communities assigned to the node. **b**. Pairwise trajectory ratio (PTR) for each of the node pairings in *a*. **c.** A static community template for the nodes in *a*. **d.** Given the static community template in *c*, the static community co-occurrence (SCC) is calculated. This is a value for each static community combination (both "within" a single static community and "between" pairwise combinations).

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

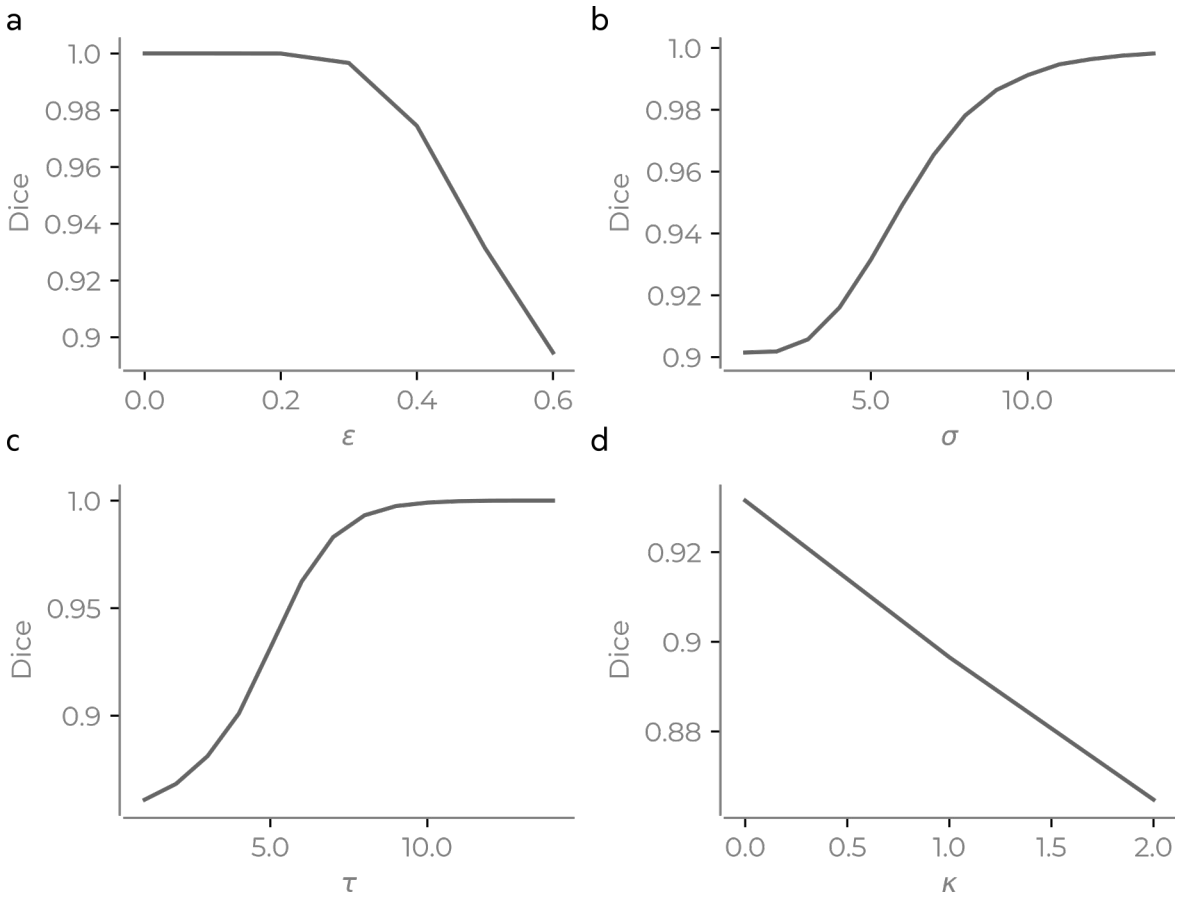

Fig. S3. Example of changing hyperparameters of TCTC for one MSC subject/session. All variables are fixed at ε = 0.5, τ = 5, σ = 5, κ = 0 except for the variable that changes in each panel. **a.** ε changes, **b.** σ changes, **c.** τ changes, **d.** κ changes. y-axis shows the average dice coefficient of each TCTC trajectories with the static community template.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

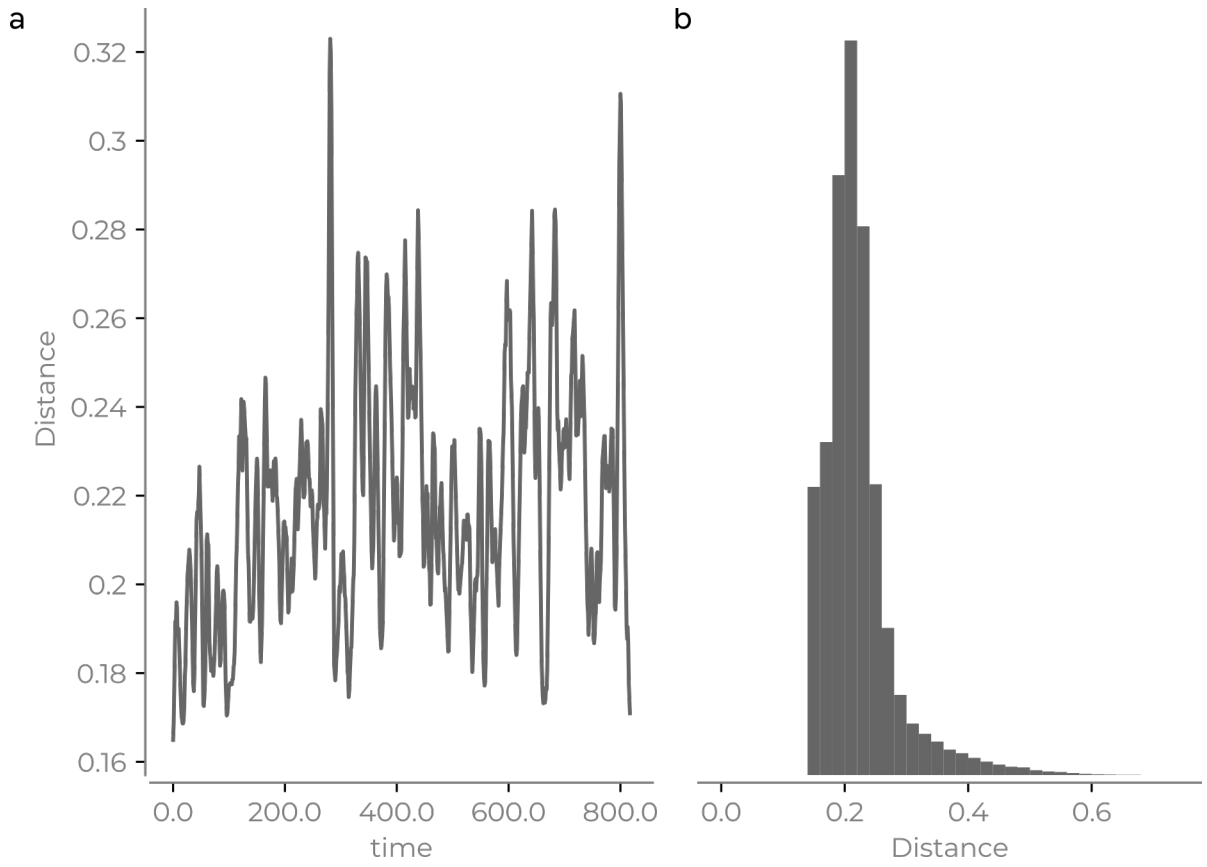

Fig. S4. Resting state temporal community fluctuations through time. **a.** Hamming distance for one example session/subject from the MSC dataset showing between the binary connectivity matrix based on the static template and the TCTC communities. **b.** Distribution of hamming distance for all subjects and sessions illustrating a heavy tailed distribution.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

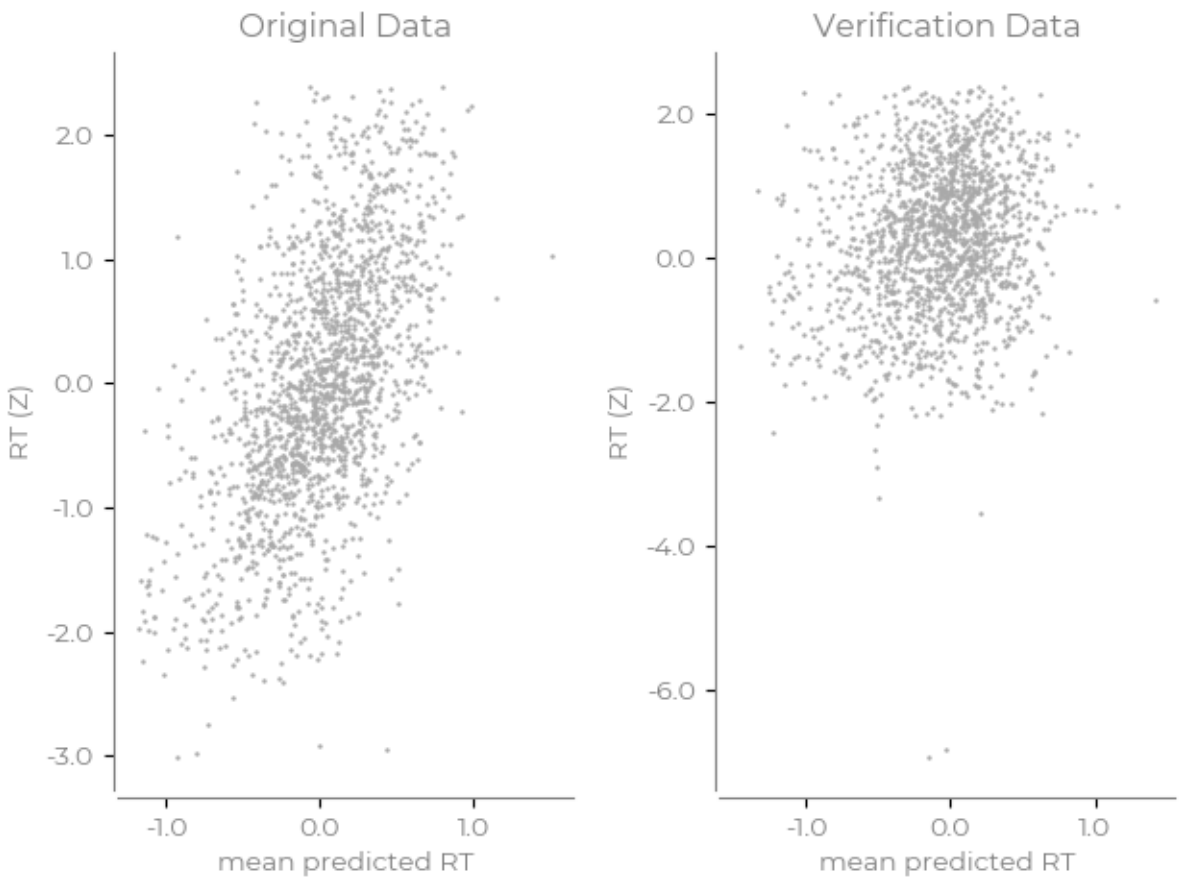

Fig. S5. Verification of RT model. Left is the average predictive posterior samples vs the corresponding outcome variable on the original data. Right is the average predictive posterior samples vs the verification data.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Tables

	loo	p_loo	d_loo	weight	se	dse
M-tcomb	4767.22	139.01	0.00	0.77	60.59	0.00
M-t1	4771.25	140.49	4.04	0.00	60.40	5.00
M-t2	4771.34	140.21	4.12	0.00	60.33	4.91
M-t0	4773.47	139.99	6.26	0.00	60.33	4.99
M-t3	4774.96	139.31	7.75	0.00	60.40	5.65
M-PC3-t1	4775.20	135.66	7.98	0.19	60.55	7.61
M-PC5-t1	4780.23	135.96	13.02	0.04	60.84	8.63
M-t4	4781.53	139.48	14.31	0.00	60.64	7.50
M-PC2-t6	4783.08	135.16	15.87	0.00	60.82	9.08
M-PC4-t8	4783.16	134.07	15.95	0.00	60.91	9.09
M-PC4-t1	4783.25	134.66	16.03	0.00	60.52	9.30
M-t5	4784.17	139.41	16.96	0.00	60.74	8.24
M-t6	4787.61	139.35	20.40	0.00	60.84	8.83
M-t7	4791.56	138.73	24.35	0.00	60.82	9.76
M-t8	4799.28	139.44	32.06	0.00	60.97	10.91
M-t9	4804.96	139.58	37.74	0.00	61.18	12.02
M-t10	4809.37	140.56	42.15	0.00	61.13	12.26

Table S1 LOO values for all models regarding reaction times. Explanation of model names: *M-tcomb* is all PC components from multiple time-points that are marked with a * in Figure 5. *M-PCX-tY* are models that are one PC component X at time-point Y. These models are the PC components marked with a * in Figure 5. *M-tX* are models that contain all 5 PC components at time-point X.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	loo	p_loo	d_loo	weight	se	dse
M-tcomb	1107.53	71.92	0.00	0.72	56.01	0.00
M-PC2-t0	1111.21	68.35	3.67	0.11	56.00	5.45
M-PC1-t10	1120.61	71.02	13.08	0.17	56.23	9.60
M-t0	1122.16	75.48	14.63	0.00	56.80	6.37
M-PC4-t10	1122.51	66.87	14.97	0.00	56.36	9.34
M-t10	1125.10	76.51	17.56	0.00	56.67	7.34
M-t1	1126.11	75.09	18.58	0.00	57.03	7.94
M-t3	1127.98	74.68	20.45	0.00	56.86	8.92
M-t2	1128.61	75.47	21.07	0.00	57.09	8.84
M-t6	1128.66	73.59	21.13	0.00	56.72	8.45
M-t5	1129.61	74.27	22.08	0.00	56.80	9.11
M-t9	1130.34	73.02	22.81	0.00	57.01	9.10
M-t4	1131.96	74.61	24.43	0.00	57.02	9.75
M-t8	1132.47	73.72	24.94	0.00	57.11	9.86
M-t7	1132.55	73.82	25.01	0.00	57.03	9.69

Table S2 LOO values for all models regarding accuracy. Explanation of model names: *M-tcomb* is all PC components from multiple time-points that are marked with a * in Figure 5. *M-PCX-tY* are models that are one PC component X at time-point Y. These models are the PC components marked with a * in Figure 5. *M-tX* are models that contain all 5 PC components at time-point X

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	loo	p_loo	d_loo	weight	se	dse
M-PC3-t7	483.70	38.33	0.00	0.57	47.74	0.00
M-t7	483.86	44.22	0.16	0.43	47.84	2.17
M-t6	485.31	42.99	1.61	0.00	47.90	1.58
M-t8	485.59	43.14	1.89	0.00	47.99	2.03
M-t9	486.20	43.64	2.51	0.00	48.04	2.23
M-t5	486.79	42.45	3.09	0.00	48.10	1.88
M-t10	487.94	42.00	4.24	0.00	48.20	1.77
M-t4	489.40	41.64	5.70	0.00	48.29	2.56
M-t3	490.54	43.56	6.84	0.00	48.61	3.54
M-t2	492.57	42.53	8.87	0.00	48.69	3.54
M-t0	496.83	43.72	13.13	0.00	49.26	4.99
M-t1	496.98	43.14	13.28	0.00	49.23	4.79

Table S3 LOO values for all models regarding response. Explanation of model names: *M-PCX-tY* are models that are one PC component X at time-point Y. These models are the PC components marked with a * in Figure 5. *M-tX* are models that contain all 5 PC components at time-point X.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PC	Behaviour	Median of posterior	Time	90% CI	% posterior above/below 0
PC2	Reaction time	0.032	6	[-0.0079, 0.073]	90.49
PC3	Reaction time	-0.072	1	[-0.11, -0.032]	99.88
PC4	Reaction time	-0.047	1	[-0.086, -0.0073]	97.42
PC4	Reaction time	-0.053	8	[-0.094, -0.013]	98.58
PC5	Reaction time	0.054	1	[0.014, 0.094]	98.66
PC1	Accuracy	0.15	10	[0.0065, 0.31]	95.70
PC2	Accuracy	0.32	0	[0.17, 0.47]	99.98
PC4	Accuracy	0.20	10	[0.056, 0.35]	98.83
PC3	Response	-0.35	7	[-0.12, -0.58]	99.34

Table S4: Description from creating new models combining credible features from __Fig. 5cde__ where peaks of the median posterior where 90% of the credible interval were above/below zero. Results describe posterior distributions in Fig. 6.

bioRxiv preprint doi: https://doi.org/10.1101/617027; this version posted April 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Methods

Walkthrough of each TCTC parameter

Let us look at an illustration how each of the different hyperparameters work and how communities are inferred from the trajectories. If *X* contains two time series, they get clustered together if they are within ε distance of each other given a distance function (__Fig. S1a__, see methods for distance functions used). However, stationary time series will often cross each other incidentally, leading to brief moments where the two unrelated time series would be incorrectly grouped in the same community. Adding a minimum time requirement, τ, (__Fig. S1b__) assures that the communities persist for a minimum number of time points. If we now add a third time series to *X*, more community structures will be possible across the time series (__Fig. S1c__). However, we may want communities to be of a minimum size; increasing the parameter σ entails that a community consists of σ number of nodes (__Fig. S1d__). Finally, it is possible that time series can contain brief bursts of uncorrelated noise (__FigS1e__). When increasing the final hyperparameter κ, it allows for some time-points to briefly violate the other rules (__Fig. S1e__). In sum, these four different hyperparameters can set the spatial and temporal resolution of the identified communities. One of the benefits of this approach is these four different parameters each have a clear interpretation of how they affect the community detection.

Disclosure of additional models run tests run

Prior to evaluating the Bayesian model for each second following stimulus onset in the hierarchical Bayesian model (Fig 5) , two models were run (“reaction time” and “response” models (the latter being slightly misspecified as it also included incorrect trials in the “response present”)) with a slightly different method to derive the network configurations. There were three differences in the initial approach:

(1) PC components were derived for only a portion of the block (1 per trial) instead of deriving the PC components across the entire block;
(2) Analysis was performed only at 7.2 seconds after stimulus onset;
(3) Only “reaction time” and “response” models were run.

Regarding (1), this resulted with only marginal differences regarding the PC components in regards to which networks had high loadings and overall variance explained but the components themselves looked visually similar. Regarding (2.) the results of these models were similar to 7 seconds after stimulus onset in Fig 5ce. PC3 (for response) and PC4 (for RT) both had over 99% of their posterior distributions below 0. Regarding (3.) the intention was to run accuracy as well but that model had not been run before the change occurred.

The motivation for changing the analysis was done after conceiving the approach performed in the main article (i.e. using the time dimension) was a more apt approach to capture the dynamics and not due to the lack of success of this initial approach and the three changes outlined above were to accommodate for this change. The verification dataset was not touched with this analysis. Only for the purpose of full transparency of exploration do we disclose this.