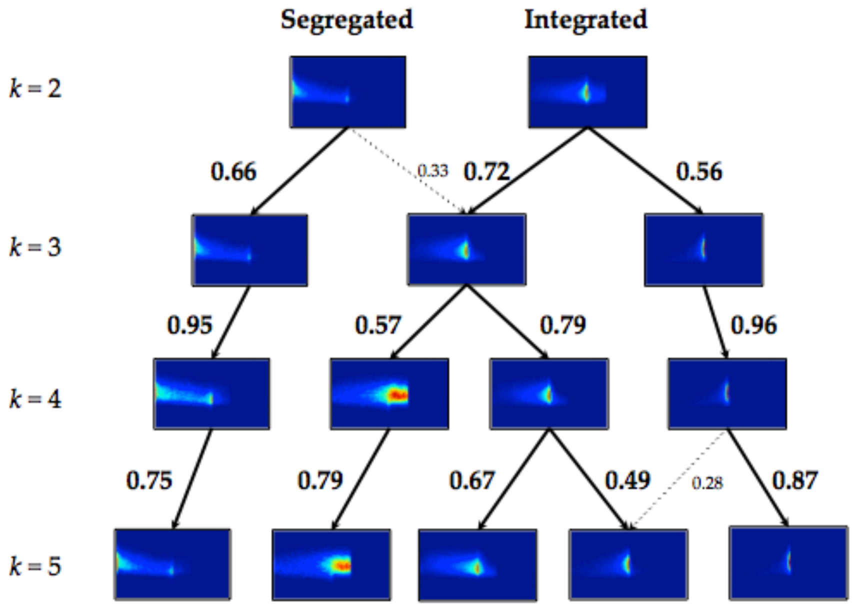

Figure S1, related to Experimental Procedures: DICE's coefficient between time series partition at k = 2 – 5 for an individual subject (HCP ID#100307) in the Discovery cohort. The values on each edge denote the Dice's coefficient between two time series – only DICE coefficients greater than 0.25 were shown in the figure. Note that, despite some minor interactions, the clustering into two states (top row) is maintained throughout lower levels of the cluster. A similar pattern was observed in each individual studied.

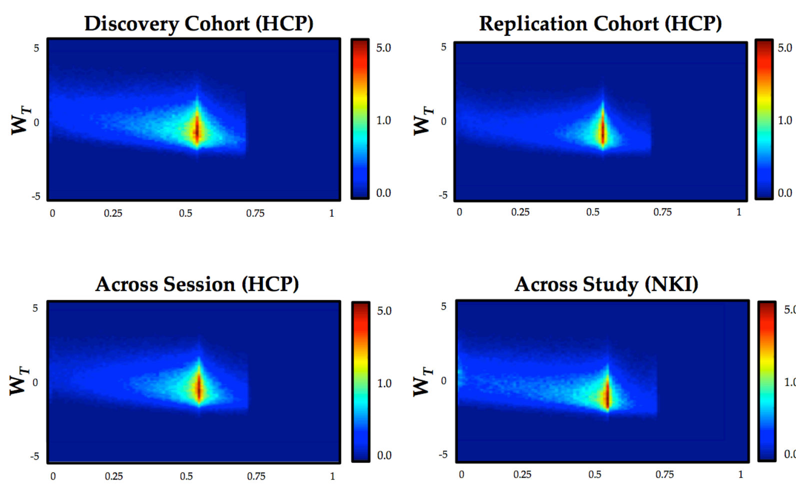

Figure S2, related to Experimental Procedures: mean joint histograms for single subject from each of the four samples using in the study, comparing the distribution of W_T and B_T values over each temporal window.

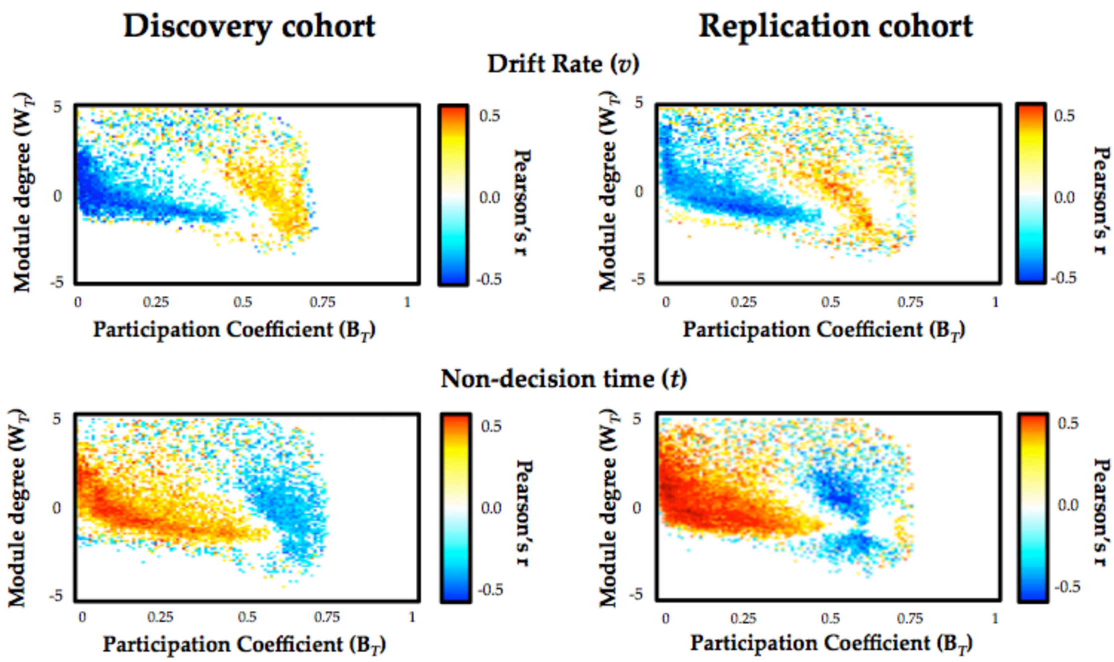

Figure S3, related to Figure 3: upper panel – group-level correlation between drift rate on the N-back task and each bin of the mean cartographic profile during the N-back task in the Discovery and Replication cohorts; lower panel – group-level correlation between non-decision time on the N-back task and each bin of the mean cartographic profile during the N-back task in the Discovery and Replication cohorts.

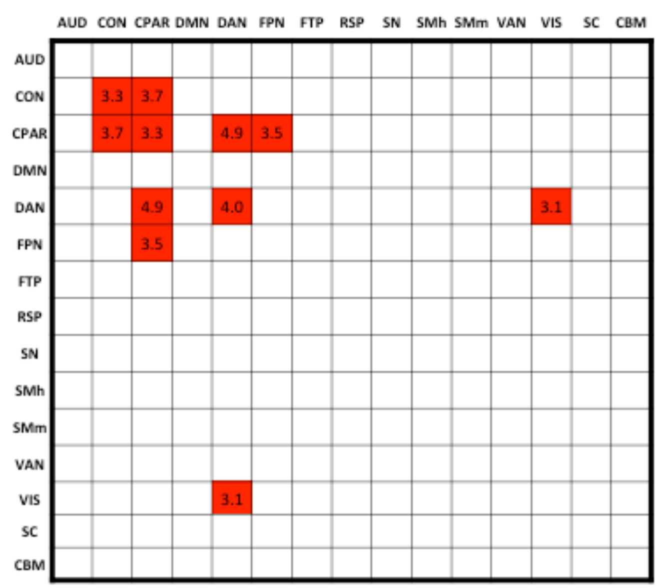

Figure S4, related to Figure 2: values within each cell designate t-statistic results of mixed effects general linear model comparing mean network connectivity in 2-back > 0-back blocks of the N-back task (FDR q = 0.05) – only values above threshold are presented. Key: AUD – auditory; CON – cingulo-opercular; CPAR – cingulo-parietal; DMN – default mode; DAN – dorsal attention; FPN – frontoparietal; FTP – frontotemporal; RSP – Retrosplenial; SN – salience; SMh – somatomotor hand; SMm – somatomotor mouth; VAN – ventral attention; VIS – visual; SC – subcortical; CBM – cerebellum.

Validation analysis

Based on the novel methodology and short temporal windows utilized in this analysis, we performed numerous additional analyses to clarify the relationship between the main temporal outcome measures calculated in this study with respect to: i) the choice of functional connectivity algorithm; ii) the use of short windows; iii) the choice of high pass filter threshold; iv) the potential adverse effects of spurious noise during the resting state scanning session; and v) the choice of community detection algorithm.

Effect of connectivity algorithm

We have previously shown that the MTD is more sensitive and specific to subtle shifts in connectivity structure than those estimated using a sliding-window Pearson's correlation in simulated BOLD data (Shine et al., 2015). To determine whether using a sliding-window Pearson's correlation gave similar results to the MTD in this study, we re-analyzed the data in our experiment using a Pearson's correlation with a sliding window of w = 14 (the window length used in our experiment) and compared the outcome measures from our study using the MTD with w = 14. Consistent with previous work (Shine et al., 2015), we found that the time-averaged estimates of the MTD were similar to those calculated using the mean Pearson's correlation matrices (W_T: spatial Pearson's r = 0.757; B_T: r = 0.902), however there were measurable differences in the mean cartographic profile (r = 0.467), particularly as it evolved over time (mean B_T: r = 0.142), suggesting that the fluctuations observed using the MTD were not the same as those estimated using a Pearson's correlation coefficient. Similar results were obtained when comparing sliding window Pearson's correlations using w = 83 (which equates to ~60 seconds, a commonly used window length in the neuroimaging literature(Leonardi and Van De Ville, 2015) with MTD with w = 83. These differences may be due to the different spectral profiles of the two connectivity measures, as the MTD preferentially focus on the highest frequency signals in the data (here, after filtering, ~0.1 Hz), whereas Pearson's correlations instead target the slowest frequency signals in the data (here, ~0.017 Hz).

Effect of window length

To determine the effects of using a short window length in the calculation of the MTD on the outcome measures in our experiment, we re-ran the analysis in all 92 subjects from the original group across a range of window lengths (10-100 in intervals of 5 TRs). We then compared the main outcomes measures in our study (parcel-wise mean W_T, mean B_T and the cartographic profile) across all window lengths in all 92 subjects using a region-wise Spearman's rho correlation. There were observable differences between the cartographic profiles when calculated at different window lengths (i.e. longer windows led to less fluctuations and hence, an estimate of the connectome as stationary). Importantly, the frequency of fluctuations (estimated as the frequency of fluctuations along the B_T axis over time) was greater than expected by a stationary null mode and also reliable at window lengths of 14 TRs (here, approximately 10 seconds), suggesting that the MTD affords a reliable means for tracking spatiotemporal dynamics in fMRI BOLD data.

Effect of high pass filter threshold

Recent work on simulated time series has suggested that low frequency signals in fMRI BOLD data can spuriously effect estimates of sliding window covariance (Leonardi and Van De Ville, 2015), and as such, it has been recommended that the lower bounds of the high pass filter used on BOLD data be set to the reciprocal of the window length used in the sliding window analysis (in our study, this would amount to a low pass filter of $\frac{1}{10.08}$ seconds; ~ 0.1 Hz). In contrast, others have shown that using data with a high signal to noise ratio can effectively mitigate the potential issues with aliasing (Zalesky and Breakspear, 2015). In addition, we have previously shown that the MTD is relatively insensitive to low frequency fluctuations (Shine et al., 2015), as the temporal differencing used to create the MTD renders the technique relatively insensitive to signals with a lower frequency than the upper bounds of the low pass filter. As such, in keeping with others (Bassett et al., 2015), we chose to use a high pass filter of 0.071 Hz in our study. However, to explicitly test whether the inclusion of low frequency altered the outcome measures in our experiment, we re-ran our analysis over a range of high pass filters (0.001 to 0.1, in steps of 0.1), with a low pass filter of 0.125. Across this entire range, we did not observe any group-level differences in our outcome measures of interest (minimum r = 0.718, p < 0.001), confirming that the signals measured by the MTD were not adversely affected by spurious low frequency signals in the data.

Effect of preprocessing strategies

To determine whether head motion was associated with each measure, we calculated the first principal component of the 12 head motion parameters over time and then correlated this new vector with each outcome measure. We also correlated the framewise displacement as well as the DVARS values with each outcome measure. None of these correlations was significant at $p < 0.05$. To test for the effects of temporal 'scrubbing' (Power et al., 2014) and global signal regression, we re-analyzed data for the original group of 92 subjects separately and then compared the outcomes measures across analyses. For the scrubbing analysis, time points associated with framewise displacement > 0.5mm or DVARS > 5% were removed and missing data points were then imputed using linear interpolation. Using this technique, we observed strong positive correlations between outcomes measures across analyses (mean Pearson's $r > 0.700$ for all comparisons), suggesting that neither preprocessing strategy had a major effect on topological or topographic measures calculated using the MTD. Finally, to determine whether physiological sources of noise were associated with each measure, we convolved the heart rate and respiratory signals with a pre-defined transfer function (Chang and Glover, 2010) and then correlated the resultant vector either each outcome measure for the 92 subjects from the discovery group. In both instances, we did not observe a significant correlation between the outcomes of interest in our analysis and physiological signals (for each comparison, $|r| < 0.050$, $p > 0.05$).

Effect of community detection algorithm

To ensure that the modularity assignment within each window was robust across community detection algorithms, we repeated the analysis using the Infomap algorithm (Rosvall and Bergstrom, 2008) in the discovery cohort and observed broadly consistent results (mean mutual information between community partitions = 0.342 ± 0.16). Furthermore, we also used another popular method, multi-slice community detection (Mucha et al., 2010), to detect community structure in the time-resolved brain. The temporal modules identified by this method (using standard parameters: $\gamma = \omega = 1$) also gave similar results (mean mutual information between community partitions = 0.447 ± 0.21). However, due to practical concerns regarding computational load and the fact that the Infomap algorithm requires the thresholding of an adjacency matrix (and as such, is insensitive to anti-correlations), we did not use either method for the full study. The fluctuations were also observed when a sub-optimal partition was used to estimate the modular assignment for each region (taken by stopping the modularity maximization approach before it reached a local minima). However, the fluctuations were not similar when a pre-defined 'hard' partition was used to estimate modular assignment, suggesting that the fluctuations represent changes in modular architecture and topology, as opposed to topology alone.

Supplemental Experimental Procedures

Experimental tasks from Human Connectome Project

Data were analyzed from the following tasks: i) a simple motor task in which the participants were presented with visual cues that required them to tap their left or right fingers, squeeze their left or right toes, or move their tongue; ii) a visually-based N-back task, which consisted of interleaved 10 second blocks of a high (2-back) and low (0-back) load N-back task, each block with object stimuli in one of four classes (places, faces, body parts and tools; iii) a social cognition task, in which subjects passively viewed videos of interacting objects and were asked to judge the character of their interactions; iv) a gambling task, which took the form of a card 'guessing' game in which subjects were rewarded for correct responses; v) a relational matching task, in which subjects were required to distinguish between items that were either related to one another conceptually or had a matching pattern; vi) a emotional processing task, in which subjects are asked to judge the emotional character of faces; and vi) a language task, in which subjects either listened to short narratives or performed a simple mathematical task.